FESTA Handbook

Version 5

Revised by FOT-Net

(Field Operational Tests Networking and Methodology Promotion)

28 February 2014

Acknowledgements

We would like to thank all the individuals who have contributed to this document.

The original Version and Version 2 were produced by the FESTA project consortium.

An update to Version 3 of the FESTA manual was performed by a working group in the FOT-Net 2 Support Action based on the results of seminars organised in the previous FOT-Net project and feedback from users of Version 2.

Version 4 took into account the feedback received after the publication of Version 3 for Open Consultation in the period July-August 2011 and during a presentation workshop in Gothenburg (8 September 2011).

Version 5 is the result of the activity of five Working Groups in the FOT-Net 2 Support Action during 2012 and 2013. The focus of Working Groups was on the following items: Data Analysis, Events and Incident Definition, Legal and Ethical Issues, Impact Assessment and Scaling Up, Data Sharing (VCC). The Working Groups proceeded in parallel, organising dedicated workshops and setting up links to running FOTs. Intermediate progress was presented to the main ITS International congress. A final workshop was then organised in Turin (4-5 November 2013) amongst experts in order to validate the content to insert and agree amendments to the new version.

The following organisations contributed directly or indirectly to one or more revisions of the FESTA handbook as members of at least one of the consortia of the FESTA, FOT-Net, and FOT-Net 2 Support Actions:

A.D.C. Automotive Distance Control Systems GmbH	Germany
Association of French Motorway Companies (ASFA)	France
BMW	Germany
Chalmers University of Technology (SAFER)	Sweden
Daimler	Germany
Delphi France	France
European Road Transport Telematics Implementation Coordination Organisation (ERTICO)	Belgium
Federal Highway Research Institute (BAST)	Germany
FIAT Research Centre (CRF)	Italy
Fraunhofer Society for the advancement of applied research	Germany
French National Institute for Transport and Safety Research (INRETS)	France
Galicia Automotive Technological Centre (CTAG)	Spain
INFOBLU SPA	Italy
International Automobile Federation (FIA)	Belgium
ITS Bretagne	France
Loughborough University	United Kingdom
Ministry of Transport, Public Works and Water Management (RWW)	Netherlands

Netherlands Organization for Applied Scientific Research (TNO)	Netherlands
Orange France	France
Promotion Of Operational Links With Integrated Services, Association Internationale (POLIS)	Belgium
Renault	France
Robert Bosch GmbH	Germany
Swedish National Road and Transport Research Institute (VTI)	Sweden
Technical Research Centre of Finland (VTT)	Finland
Technical University of Aachen (IKA)	Germany
University of Koln	Germany
University of Leeds	United Kingdom
Volvo Car Corporation	Sweden
Volvo Technology Corporation	Sweden

Special thanks are due to the various external experts for their valuable insights from their large, experience and contributions in conducting FOTs. They contributed to the original version produced by the FESTA consortium. These experts include Jim Sayer of UMTRI, Charlie Klauer of VTTI, Tom Triggs, Kristie Young, Eve Mitsopoulos-Rubens, Nebojsa Tomasevic and Karen Stephan of MUARC, Harri Peltola of VTT and Riku Kotiranta of Chalmers University of Technology.

Disclaimer:

The FESTA and FOT-Net 2 Support Actions have been funded by the European Commission DG Information Society and Media in the 7th Framework Programme. The content of this publication is the sole responsibility of the project partners listed herein and does not necessarily represent the view of the European Commission or its services.

List of Abbreviations[1]

ABA	Aggregation Based Analysis
AC	Alternating Current
ACC	Adaptive Cruise Control
ACEA	European Automobile Manufacturers' Association
ADAS	Advanced Driver Assistance Systems
BCR	Benefit Cost Ratio
BLOB	Binary Large Object
C2C-CC	Car-to-Car Communication Consortium
CALM	Continuous Air interface for Long and Medium range communication
CAM	Cooperative Awareness Message
CAN	Controller Area Network
CBA	Cost Benefit Analysis
CEN	Comité Européen de Normalisation (European Committee for Standardization)
CPI	Consumer Price Index
CPU	Central Processing Unit
CRE	Crash Relevant Events
DAQ	Data Acquisition
DAS	Data Acquisition System
DBA	Database Administrator
DC	Direct Current
DENM	Decentralized Environmental Notification Message
DGPS	Differential Global Positioning System
EBA	Event Based Analysis
ECU	Electronic Control Unit
EDR	Event Data Recorder
EMC	Electromagnetic Compatibility
ETSI	European Telecommunications Standards Institute
EuroNCAP	European New Car Assessment Programme
FCW	Forward Collision Warning
FESTA	Field opErational teSt supporT Action
FFM	Five Factor Model
FM	Frequency Modulation
FOT	Field Operational Test
GDP	Gross domestic product
GIS	Geographical Information System
GMaps	Google Maps
GMM	Gaussian Mixture Models
GNSS	Global Navigation Satellite System
GPS	Global Positioning System

[1] When relevant an Hyperlink to Web is added

GSM	Global System for Mobile communications
GUI	Graphical User Interface
HMI	Human Machine Interface
I2V	Infrastructure to Vehicle(see Vehicular communication systems)
ICT	Information Communication Technology
IEEE	Institute of Electrical & Electronics Engineers
I/O	Input/Output
IRR	Internal Rate of Return
ISA	Intelligent Speed Adaptation
ITS	Intelligent Transportation System
LDW	Lane Departure Warning
LED	Light Emitting Diode
LIDAR	Light Detection and Rating
LIN	Local Interconnect Network
MJPEG	Motion JPEG
MOST	Media Oriented Systems Transport
MP3	Moving Picture Experts Group Layer-3 Audio
MPEG	Moving Picture Experts Group
MPEG-4	MPEG standard, version 4
NAS	Network Attached Storage
ND	Nomadic Device (often referred as Mobile Device)
NDD	Naturalistic Driving Data
NDS	Naturalistic Driving Study
NHTSA	National Highway Traffic Safety Administration
NPV	Net Present Value
NTP	Network Time Protocol
OEM	Original Equipment Manufacturer (in this document used as synonym of vehicle manufacturers)
OSM	Open Street Map
PI	Performance Indicator
PND	Personal Navigation Devices
RAID	Redundant Arrays of Independent Disks
RR	Range Rate
RSU	Road Side Unit (see Vehicular communication systems)
SAN	Storage Area Network
SMS	Short Message Service
SPSS	Statistical Package for the Social Sciences
SQL	Structured Query Language
SSD	Solid State Drive
SSS	Sensation Seeking Scale
T-LOC	Traffic Locus of Control
TTC	Time To Collision (relevant parameter in preventive safety applications – see e.g. Prevent)
USB	Universal Serial Bus
UTC	Universal Time Coordinated
V2I	Vehicle to Infrastructure (see Vehicular communication systems)
V2V	Vehicle to Vehicle (see Vehicular communication systems)

V2X	Vehicle to X (Any Station) (see Vehicular communication systems)
VMS	Variable Message Sign
WBS	Work Breakdown Structure
WLAN	Wireless Local Area Network
WTP	Willingness To Pay
WYLFIWYF	What-You-Look-For-Is-What-You-Find
ZFS	Zettabyte File System

Table of Contents

Table of Contents iv

1. Introduction 1

   1.1 The Naturalistic Driving Study in relationship to the Field Operational Test and the FESTA methodology 4

2. Planning and Running a Field Operational Test 8

   2.1 Introduction 8

   2.2 The FOT Implementation Plan 8

3. Legal and Ethical Issues 14

   3.1 Introduction 14

   3.2 Participant recruitment 15

   3.3 Participant agreement 15

   3.4 Data protection and data ownership 16

   3.5 Risk assessment 17

   3.6 System safety 18

   3.7 Approval for on-road use 18

   3.8 Insurance 18

   3.9 Responsibilities 18

   3.10 Video data collection (specific issues) 19

   3.11 Ethical approval 19

   3.12 Iteration 19

4. From Functions to Hypotheses 20

   4.1 Systems and functions 21

   4.2 General methodology 29

5. Performance Indicators 45

   5.1 Introduction 45

   5.2 Performance indicators definition 45

   5.3 Measures 46

   5.4 The PI-Measures-Sensors matrix 53

   5.5 Performance Indicators per impact area 53

   5.6 Iteration 60

6. Experimental procedures 62

6.1 Participants 62
6.2 Study design 65
6.3 Experimental environment 69
6.4 Conducting a pilot study to test the evaluation process 75
6.5 Controlled testing 77
7. Guidelines for Data Acquisition 83
7.1 The measures and sensors tables 84
7.2 Data acquisition 84
7.3 Specific sensors 87
7.4 Mechanical requirements 92
7.5 Electrical requirements 93
7.6 DAS data storage 94
7.7 System configuration 98
7.8 Acquisition of data 99
7.9 Synchronisation 99
7.10 DAS status and malfunction management 101
7.11 System installation 102
7.12 Proprietary data 103
7.13 Personal integrity and privacy issues in data acquisition and analysis 103
7.14 On-line quality management procedures 104
8. Guidelines for - Databases and Analysis Tools 106
8.1 Database design and implementation 106
8.2 Database and data storage implementation 109
8.3 Off-line quality management procedures 111
8.4 Data analysis tools 113
8.5 Database usage 114
9. Data Analysis and Modelling 115
9.1 Introduction 115
9.2 Large Data-set handling 116
9.3 Consistency of the chain of data treatment 118
9.4 Precision in sampling 119
9.5 Requirements for integration and scaling up 120
9.6 Appropriate techniques at the five links of data analysis 123
10. Impact Assessment and Socio-Economic Cost Benefit Analysis 129

10.1 Introduction. 129
10.2 Considerations 130
10.3 Analysis of impacts 133
10.4 Socio-economic Cost Benefit Analysis 140
11. References 156
List of Annexes 163

1. Introduction

Field Operational Tests (FOTs[FW])[2] were introduced several years ago as an evaluation method for driver support systems[FW][3] and functions[FW][4] with the aim of proving that such systems can deliver real-world benefits. . A number of such FOTs have been conducted at a regional, national and international (European) level, to evaluate a range of systems, particularly a variety of driver support systems. These FOTs have proven to be highly valuable and have been identified as an important means of verifying the real-world impacts of new systems, including at a European level and as a means to verify that previously conducted R&D has the potential to deliver identifiable benefits. The key deliverable for "**FESTA**", (Field opErational teSt supporT Action), has been to produce comprehensive guidance to facilitate the successful delivery of FOTs. This, the latest version of the FESTA handbook has drawn from experience and knowledge across numerous FOTs in order to revise and update the best practice presented here. This Handbook is the result of a joint effort of several research institutes, OEMs and other stakeholders from across Europe to prepare a common methodology for European FOTs. It is also highly relevant, and it is hoped useful, for FOTs conducted at a regional or national level within Europe as well as outside Europe.

For the purposes of this Handbook, a "Field Operational Test" (FOT) is defined as:

> A study undertaken to evaluate a function, or functions, under normal operating conditions in road traffic environments typically encountered by the participants using study design so as to identify real world effects and benefits.

This means that it must be possible to compare the effects that the function has on traffic with a baseline condition during which the function is not operating. In order to achieve this, the participants' control over or interaction with the function(s) has to be manipulated by the research team. "Normal operating conditions" implies that the participants use the platforms during their daily routines, that data logging works autonomously and that the participants do not receive special instructions about how and where to drive. Except for some specific occasions, there is no experimenter in the vehicle, and typically the study period extends over at least a number of weeks. (FOTs must however consider whether it is still possible to achieve a baseline where all the studied systems are off due to the proliferation of functions in standard vehicles).

[2] Key items have internal links and, when avaialble, a link to FOT-Net Wiki glossary, indicated as [FW]

[3] A system is defined as "a combination of hardware and software enabling one or more functions".

[4] A function is defined as "an implementation of a set of rules to achieve a specified goal".

The main purpose of this Handbook is to provide guidelines for conduct of FOTs and NDS (Naturalistic Driving Study) although, for simplicity, in the following text, the name FOT is used and differences between the two kinds of studies are summarised when relevant. It walks the reader through the whole process of planning, preparing, executing, analysing and reporting an FOT and it gives information about aspects that are especially relevant for a study of this magnitude, such as administrative, logistic, legal and ethical issues. Another aspect of the Handbook is to pave the road for standardisation of some aspects of FOTs, which would be helpful for cross-FOT comparisons. It has to be kept in mind, though, that many traffic parameters in different European countries differ substantially.

In Figure 1.1 the steps that need to be carried out during an FOT are presented in the form of V diagram, where there is correspondence between the levels on the left-hand and right-hand sides. The steps will be explained in detail in the different chapters of the Handbook. For orientation purposes, a copy of the figure is provided in the beginning of each chapter highlighting which step of the FOT Chain is described in the current chapter. The FOT Implementation Plan takes up all the steps and integrates them into one big table which can be used as a reference when actually carrying out an FOT.

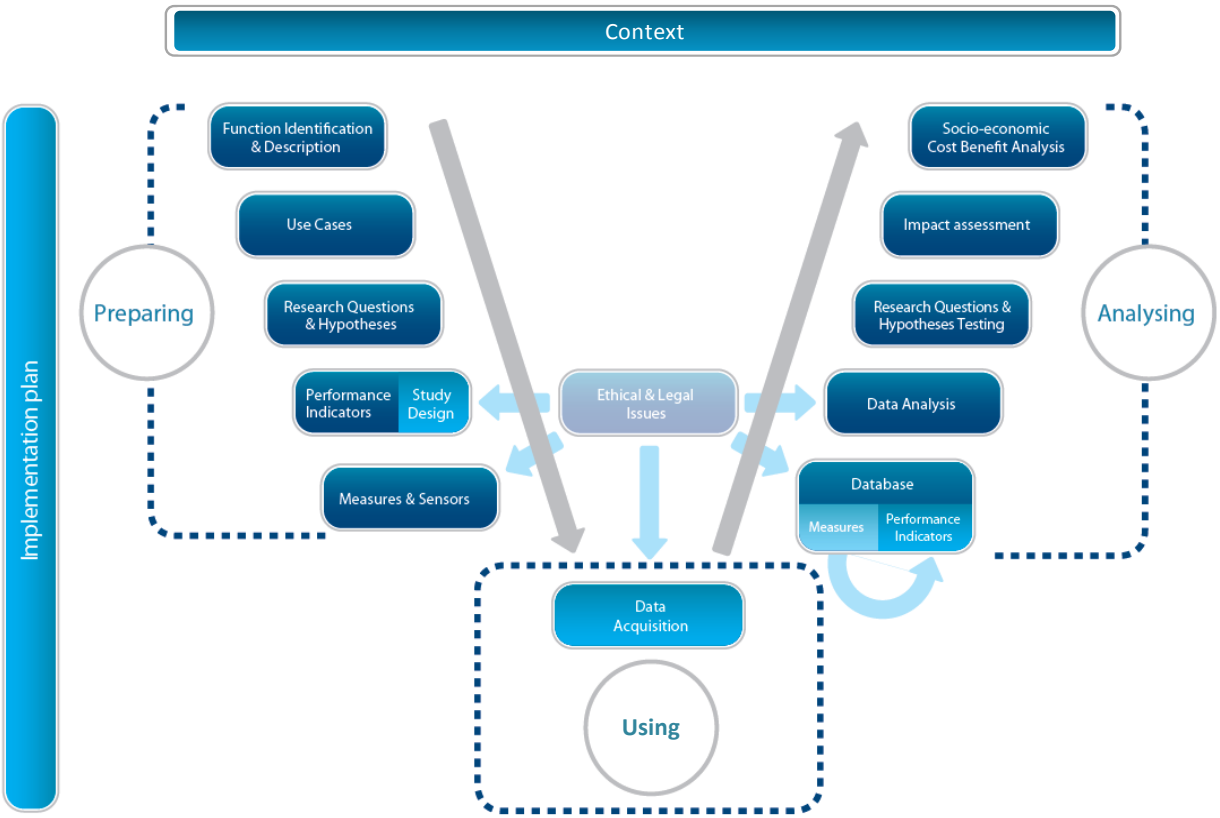


Figure 1.1: The steps that typically have to be considered when conducting an FOT. The large arrows indicate the time line.

In order to make the picture more complete a horizontal bar has been added on top of the diagram that in principle summarises the context in which the FOT is supposed to take place. For instance, the choice of a function to be tested implies that there is either a problem that is to be addressed and that the chosen function is defined to solve the problem

or that a policy objective is stated and that the function tested can be used to reach the objective. An FOT can always be related to a wider perspective than is defined by just a description of the function to be tested.

The top of the V covers setting up a goal for the study and selecting a suitable research team, and also the last steps that include an overall analysis of the systems and functions tested and the socio-economic impact assessment, dealing with the more general aspects of an FOT and with aggregation of the results. The further down on the FOT Chain V-Shape the steps are located, the more they focus on aspects with a high level of detail, such as which performance indicators[FW] to choose, or how to store the data in a database. The ethical and legal issues have the strongest impact on those aspects, where the actual contact with the participants the data handling and potential data sharing takes place.

The representation of the FESTA methodology in the form of a V does not mean that designing and performing an FOT is always a linear process. Decisions made at a certain stage of the FESTA V influence the next steps and it is likely that there will be a need to sometimes go back and redo some steps. Especially on the left-hand side of the V iteration may be necessary. For example, one may find out that the measures and sensors available do not make it possible to investigate the hypotheses[FW] defined earlier, so adjustments to the hypotheses or performance indicators may be needed. Also the right hand side of the FESTA V may influence the decisions to be made on the left-hand side. The need to assess of the socio-economic impact may influence the definition of functions, use-cases, research questions and other elements of the left-hand side. Consideration of the resources available for data analysis may also lead to revision of the left-hand side. There is, however, the question “when does one stop the iterative process?” From a research perspective, this is a continuous process. However, from a project management perspective, boundaries have to be set to reflect budget constraints and timing aspects.

The first step in the FESTA V is the identification of functions to be tested. Sometimes this may not be the best step to start with. For example, an FOT may not be driven by the technical systems that need to be tested but by a research question or an impact area. When there is a large set of functions available from which a few need to be selected as candidates for testing, definition of the research questions may help the selection process. For example if safety is to be investigated, different functions may have a higher priority than when traffic efficiency is the main focus.

In conclusion, the FESTA V provides a static picture of the complex design and execution of an FOT but in reality a more iterative process will be needed, with a starting point suited to the specific aims of a project.

The FESTA Handbook is not meant to be a substitute for consultations with experts, organising a good and capable research team, and carrying out specific investigations into the legal and ethical issues that apply to the current question and situation. It is not an exhaustive action list, and each FOT has its own special issues and concerns that have to be dealt with on an individual basis. Nor is the advice in it necessarily perfect and representative of the state of the art. On many issues, there will be scope for disagreement with the recommendations or use of alternative sources of advice. But it would certainly be

preferable for major departures from the advice to be justified to funding agencies and major stakeholders.

The FESTA project consortium decided early in the project that the primary focus of the FESTA Handbook would be on the evaluation of Advanced Driver Assistance Systems (ADAS) and In-Vehicle Information Systems for vehicles — both in the form of autonomous systems and of Cooperative Systems. It was also agreed that the FESTA Handbook should be relevant to the evaluation of Original Equipment Manufacturer (OEM), aftermarket and nomadic systems. The Handbook is therefore designed specifically to guide the evaluation of impacts of such systems, and is less relevant to the evaluation of electronic road infrastructure such as Variable Message Signs (VMS). However, it will be seen that many of the activities identified in the Handbook are common to the evaluation of most vehicle- and infrastructure-based ICT technologies).

In conclusion, the FESTA Handbook gives an overview and general guidelines concerning the conduct of an FOT. FOTS are designed to contribute to the identification and verification of solutions to a problem, and this handbook is intended to provide a formalised and practical framework, and not a cook book: the methodology described will necessarily have to be adapted to the specific case, in order to increase the efficiency of the approach or to tackle data incompleteness or inconsistency. Furthermore, the results of an FOT may have to be integrated with external sources of information, to achieve a wider perspective, and an increased relevance for tackling the problem at hand.

In addition to the Handbook itself, more detailed work, which was produced during the FESTA project and later in the FOT-Net projects, and which is referenced in the Handbook, is included in the annexes to the Handbook and the FESTA deliverables.

1.1 The Naturalistic Driving Study in relationship to the Field Operational Test and the FESTA methodology

1.1.1 Definition

The Naturalistic Driving Study (NDS) or observation is a relatively new research method using advanced technology for in-vehicle unobtrusive recording of driver (or rider) behaviour during ordinary driving in everyday traffic situations. This method yields unprecedented knowledge primarily related to road safety, but also to environmentally friendly driving/riding and to traffic management. A central focus of naturalistic driving studies is to understand explanatory factors associated with crashes and predict involvement in crashes. The naturalistic method "refers to a method of observation that captures driver behaviour in a way that does not interfere with the various influences that govern those behaviours." (Boyle et al., 2009). Naturalistic driving studies are defined as "those undertaken using unobtrusive observation or with observation taking place in a natural setting" (Dingus et al., 2006). The following characteristics have been chosen to define the Naturalistic Driving approach in the recent EU project PROLOGUE (Sagberg et al, 2011):

- Unobtrusive recording of driver and vehicle parameters

- Normal driving, i.e. driving purpose and driving destinations as defined by the driver, and driving taking place on roads open to ordinary traffic, and with the vehicle that the driver normally uses (owned, leased, or company vehicle)
- No observer present in the vehicle

1.1.2 Relation to FOTs

Naturalistic Driving Studies tend to focus on crash-explanatory factors, and Field Operational Tests generally focus on evaluation of systems or functions. However the collected data in both types of studies can be used for many alternative purposes such as analysis of Environment, Efficiency and Mobility impacts. NDS and FOTs are preferably seen as different methods because (a) the study design is different (participant selection, experimental conditions, vehicle sample, etc.), and (b) the research questions and hypotheses are different. In particular, the main difference relates to degree of experimental control found in the study, as illustrated in Figure 1.2. It is recognized that Naturalistic Driving Studies and Field Operational Tests have some common methodological aspects and that there are gradual transitions between FOTs and NDS.

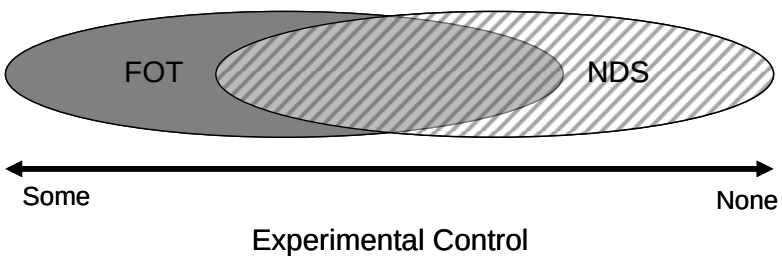


Figure 1.2: A partially overlapping relationship between Field Operational Tests and Naturalistic Driving Studies along a continuum of experimental control.

A more generic methodological approach is emerging whereby naturalistic data collected by either NDS or FOTs can be used for similar purposes. For example, an NDS may be used to evaluate the impact of on-market intelligent safety functions (Antin et al., 2011) and an FOT may be used to study crash causation (Olson et al., 2009). Even though there may be differences in purpose between NDS and FOT studies, the technology for driver behaviour observation may be the same, and consequently experiences from FOTs have been important inputs to the planning of a large-scale European NDS (Sagberg et al. 2011). Between the FOT and the NDS, lies the Naturalistic FOT. The Naturalistic Field Operational Test (N-FOT) is defined as a study undertaken using unobtrusive observation in a natural setting, typically to evaluate the relationship between (permanent or temporary) driver-, vehicle-, or environment factors with crash risk, driving behaviour, and countermeasure effectiveness (Victor et al. 2010). This definition accommodates for both accident[5] research oriented research of explanatory factors associated with crashes (common in NDS), as well

[5] An accident is defined in this handbook as an undesired event that results in personal injury or property damage.

as for the evaluation and development-oriented research on new technology and solutions (common in FOTs). N-FOT studies can for example, assess the relationship of an in-vehicle system (dynamic vehicle factor) or age (static driver factor) or distraction (dynamic driver factor) or speed cameras (static environment factor) with crash-risk, driving behaviour and/or countermeasure effectiveness. Environment sensing and video are believed to be essential for identifying near collisions and other incidents[6], and for validating that intelligent vehicle systems (e.g. collision warning, lane departure warning and intelligent speed adaptation) perform as expected.

1.1.3 NDS and the FESTA V

Although the FESTA V was originally developed as an implementation planning tool for FOT studies, it is a highly relevant and useful tool also for naturalistic driving studies. However, some modifications are required. These modifications are primarily related to the shift in focus away from function assessment in FOTs towards *behaviour* assessment in NDS. Broadly speaking, the bottom or tip of the FESTA V is most relevant for NDS and the top of the FESTA V is less directly relevant. Figure 1.3 Illustrates how the FESTA V is modified by removing four steps which are more relevant for FOTs.

The horizontal Context bar (Figure 1.3) is also useful for the NDS, it is recommended that activities which are dealing with the more general aspects of an NDS and with high level aggregation of the results take place within this horizontal Context bar. According to Sagberg et al (2011) this type of activity includes:

- User/Stakeholder Identification,
- Topics Selection,
- Dissemination, and
- Identification of Constraints, such as available technologies and budgets.

The FESTA V has two "scaling up" steps : Impact Assessment and Socioeconomic Cost Benefit Analysis. Naturalistic driving studies also have scaling up activities, such as aggregation of research questions results, and analysis of the implications of results. For example:

- Provide an integrated overview of the NDS findings.
- Use findings to identify new and more efficient safety and sustainability measures related to vehicles, drivers and road infrastructure (such as incentives to encourage adoption of new technologies, education of drivers, regulation enforcement).
- Identify ways for measures and tools to improve safety and sustainability of road transport in Europe based on NDS results.
- Demonstrate how Naturalistic Driving can be used in industrial development of safety and sustainability functions and services.

Although it can be debated whether these activities should be carried out in a separate step or box in the FESTA V, it is suggested that these activities should take place in the research questions *and* hypotheses *testing* step and the horizontal *Context* bar.

[6] An incident is defined in this handbook as an unplanned, undesired situation which has the potential to cause harm to a person or property damage.

For further detailed examples of how the FESTA V can be interpreted for Naturalistic Driving Study purposes see Sagberg et al. (2011) and Victor et al. (2010).

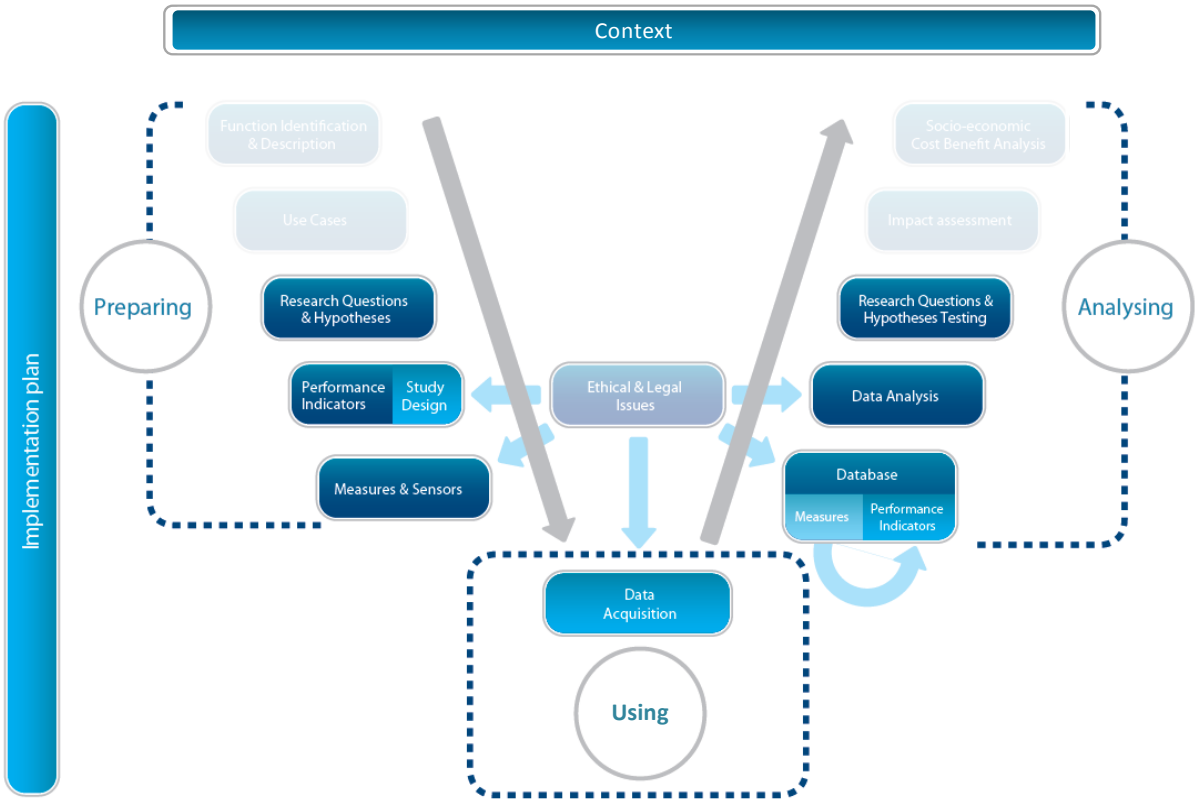
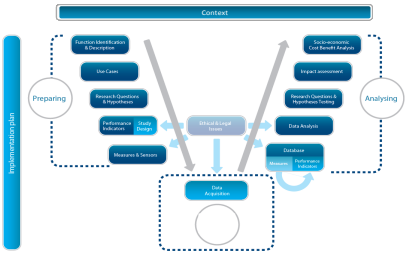


Figure 1.3: Modification of the FESTA V to match Naturalistic Driving Study purposes.

2. Planning and Running a Field Operational Test

2.1 Introduction



For a Field Operation Test(FOT[FW]) to proceed smoothly, a plan of action must be developed which documents the scientific, technical, administrative and procedural activities and tasks that are needed to successfully complete it. Given that the lifecycle of an FOT typically evolves through many phases, there are many issues to consider. In this chapter, the critical activities and tasks which are necessary to run a successful FOT are documented — in the form of a "FOT Implementation Plan" (FOTIP) — drawing on lessons learned from previous FOTs conducted in the different areas of the world.

The FOTIP is contained in Annex B of the FESTA Handbook. In this chapter, the FOTIP is introduced, described, explained and discussed.

2.2 The FOT Implementation Plan

2.2.1 Purpose

The FOTIP is intended to serve primarily as a checklist for planning and running FOTs:

- to highlight the main Activities and Tasks that would normally be undertaken in successfully completing an FOT;
- to ensure that, in running an FOT, researchers and support teams are aware of critical issues that influence the success of the FOT; and
- by drawing on the experiences of previous FOTs, to highlight the "dos" and "don'ts" of running an FOT;
- to provide a consistent framework for planning, running and decommissioning FOTs.

The FOTIP presented in this Handbook is not intended to be prescriptive, but rather to serve as a generic guide in conducting FOTs. By their very nature FOTs are major projects – extensive and expensive. Significant previous FOTs that have not delivered their anticipated outcomes have not done so primarily because of failures to anticipate problems that compromised their successful execution. The FOTIP attempts to map out all known critical issues that need to be taken into account in planning and undertaking an FOT.

The history of FOTs suggests that no two will be the same, and that there often are many unforeseen Tasks and Sub-Tasks that arise during its lifecycle. The list of Tasks and Sub-Tasks contained in the FOTIP in Annex B of this Handbook is not, therefore, exhaustive. It is based on the collective wisdom of those that have been involved in planning and running previous FOTs. There may be specific requirements for future FOTs conducted in Europe that will need to be decided on a case-by-case basis, especially where FOTs are extended to cover Naturalistic Driving Studies and potentially, other transport modes such as public transport, cyling, walking or multi-modal studies.

The FOTIP at Annex B describes *what* needs to be done, and approximately *when*, in running a successful FOT. Other relevant chapters in the FESTA handbook describe in detail *why* these activities are necessary and *how* they are to be accomplished.

2.2.2 Description of the FOT Implementation Plan

The FOTIP in Annex B of this Handbook resembles a traditional Work Breakdown Structure (WBS), but without timelines. It is specifically designed in this way so that timelines can be inserted at a later date by those responsible for the overall planning and running of the FOT.

The FOTIP is divided into three columns and two sections below each activity:

- **Column 1 — Activities**. An Activity is a high level task e.g. "Convene FOT research and support teams" that is usually needed to run an FOT.
- **Column 2 — Tasks and Sub-Tasks**. A Task directly supports an Activity e.g. "Appoint FOT project manager". A Sub-Task directly supports a Task. Essentially, this column contains a series of action statements – "do this"; "do that"; etc. There are very few sub-tasks listed in this column, to contain the size of the document. The document is cross-referenced to other chapters of the FESTA Handbook, which identify the relevant Sub-Tasks that support these Tasks.
- **Column 3 — Person/Organisation Responsible for Activity**. This column identifies the person, team, organisation or combination thereof that would usually be responsible for completion of a Task. The FOT project manager is ultimately accountable for successful completion of all Tasks, and is therefore included for every Task.**After section 1 – Critical Considerations (the "dos" and "don'ts").** This column contains critical advice for ensuring that an Activity or Task is successfully completed. e.g. "Be sure that the vehicle systems are designed so they do not drain the battery when the vehicle engine is not running." e.g. "Do not underestimate the amount of time required to recruit company drivers for the FOT."
- **After section 2 – General Advice.** This column provides general advice on how to maximise the likelihood of running a successful FOT e.g. "The FOT lifecycle is long. Hence, it is advisable to write separate reports on each critical stage of the FOT." This column also contains explanatory notes, reference to other relevant documents (e.g., FOT reports) and cross-referencing to other chapters in the FESTA Handbook.

The Activities and Tasks identified in the FOTIP are consistent with those identified in the higher level "FOT Chain" that is described in Chapter 1 of this Handbook (see Figure 1.1), although the chronological order in which the Activities and Tasks are shown varies slightly between the two. For example, in the FOT Chain, in Figure 1.1, it is assumed that the first step when planning an FOT is the identification of systems and functions to be analysed. In the FOTIP, on the other hand, this task is identified later in the sequence of planning activities (within Activity 2), as there are other planning activities and tasks that necessarily precede the identification of systems and functions to be analysed. The FOTIP identifies the scientific, technical, administrative and procedural activities for planning and running an FOT

the FOT chain summarizes the key, high level, scientific and technical steps undertaken when performing an FOT, and the sequential links between them.

Decisions taken at earlier stages may have a big impact on the FOT development and the analysis phase. Based on this concept, involving since the beginning all experts that will be encharged of the activities in the successive phases, is theoretically the safest strategy. But this is not economically sustainable and also will create large teams analysing each issue, with the results of expanding timing. For this reason a key person is the FOT project manager. He should ensure that sign-off of different activities is properly validated consulting or involving in specific meetings the right people. Although this preventive policy is applied, criticalities depending on decisions taken in previous activities/tasks may happen. In these cases negative or partially negative sign-offs may result and re-analysis of previous activities or tasks should be done. **Figure 2-1** illustrates this potentially cyclic process.

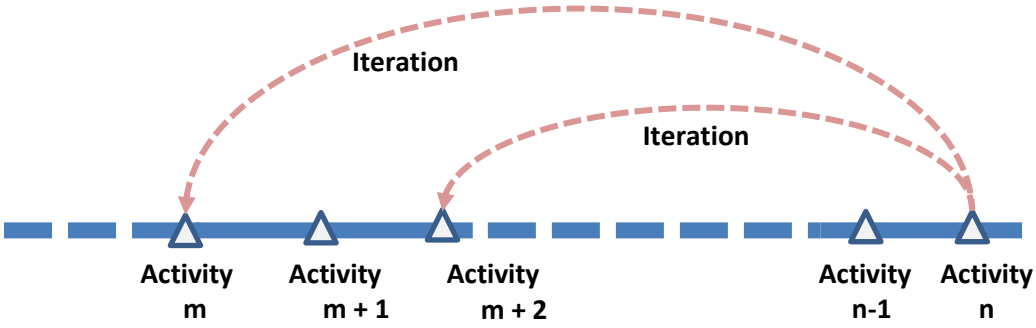


Figure 2-1 – Iterative development process

2.2.3 Development of FOT Implementation Plan

The content of the FOT Implementation Plan derives originally from several research activities undertaken in the FESTA project, based on comprehensive review of the related literature and several consultations (via worskshop, teleconference, written feedbacks) of experts member of the consortium and external experts who previously conducted FOTs. It was then revised in rev 4 and rev 5 based on the feedback received in FOT-Net and FOT-Net 2.

2.2.4 Assumptions underlying the FOT Implementation Plan

There is more than one way to perform a successful FOT . The review of the literature on FOTs revealed that many different approaches have been taken in planning, running, analysing and decommissioning FOTs. The FOTIP in Annex B of this Handbook draws together procedural activities that are most common to the known FOTs that have been conducted, and the collective wisdom of those who conducted them.

The FOT Implementation Plan is relevant to FOTs in which the ADAS and In Vehicle Information systems to be evaluated are already available on the market, or to studies in

which the systems to be evaluated must be chosen by the FOT project team, purchased or developed, and installed (e.g., as in Regan et al., 2006).

The FOT Implementation Plan provides only a general guide to the sequence in which the required activities should be performed. Some need to happen early in the project and others at the end. Some need to immediately precede others. Other tasks need to proceed concurrently with others. Decisions about the scheduling of Activities, Tasks and Sub-Tasks are the responsibility of the FOT Project Manager. Table 2-1 lists the 22 Activities identified in the FOTIP, and highlights the main dependencies that exist between them. Within Activities, it is up to the FOT Project manager to further decide which Tasks and Sub-Tasks should proceed sequentially and in parallel.

Some of the major Tasks listed in the FOTIP (e.g. “recruit participants”, within the Activity “Run FOT”) are given only a one-line description and, as such, may appear to be down played in the plan. A judgement had to be made about how much detail to include in the FOTIP. Where such one-liners exist, this is because either the Task in question is one that most researchers would normally be familiar with (e.g., recruiting study participants) or because the Sub-Tasks involved are described in detail in other Chapters of the FESTA Handbook. Where appropriate, any known difficulties and concerns associated with major Tasks for which only a one-line description is given are emphasised.

Table 2-1: A generic guide to scheduling the 22 Activities described in the FOTIP in Annex B of the FESTA Handbook.

	Set Up/Design		Preparation			Data Collection			Completion	
Convene teams and people										
Define aims, objectives, research questions & hypotheses										
Develop project management plan										
Implement procedures and protocols for communicating with stakeholders										
Design the study										
Identify and resolve legal and ethical issues										
Select and obtain FOT test platforms (vehicles, mobile devices, road side units, ...)										
Select and obtain systems and functions to be evaluated										
Select and obtain data collection and transfer systems										
Select and obtain support systems for FOT platforms										
Equip FOT test platforms with all systems										
Implement driver feedback and reporting systems										
Select / implement relational database for storing FOT data										
Test all systems against functional requirements and performance specifications										
Develop recruitment strategy and materials										
Develop driver training and briefing materials										
Pilot test FOT equipment, methods and procedures										
Run the FOT										
Analyse FOT data										
Write minutes and reports										
Disseminate the FOT findings										
Decommission the FOT										

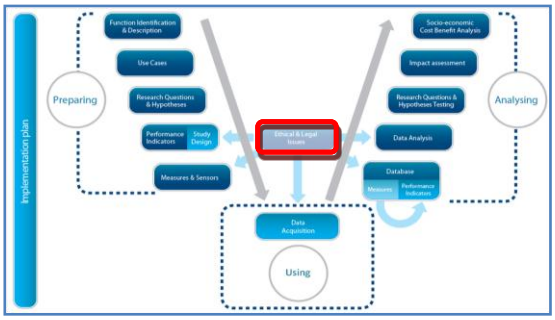
2.2.5 Using the FOT Implementation Plan

It is suggested that the FOTIP be used as follows:

- read through the FOTIP before starting to plan an FOT;

- use the FOTIP as a checklist for guiding the planning, design and running of the FOT — and as a quality control mechanism for ensuring during the study that nothing critical has been forgotten;
- read the FOTIP in conjunction with other chapters in the FESTA Handbook, and refer to other chapters and other FOT reports for detail; and
- if desired, use the FOTIP as the basis for the development of GANTT charts and other project management tools.

3. Legal and Ethical Issues



3.1 Introduction

Carrying out an FOT[FW] usually means to ask participants to give insight in their mobility behaviour. Although participants may be given some form of (financial) compensation, or in the case of fleet drivers it may form part of their job, usually participants join in order to help advance road-safety, mobility and innovation. An FOT project should therefore value their participants and treat their data and privacy with respect. Not only legal rules should be followed, but also ethical principles, defined as respect for the person[ality] and his or her autonomy, dignity and self-determination, should be a major guideline for conducting the FOTs. Risks for participants should be minimalised as much as possible.

Next to care for participants, the rights of non-participants, affected by the FOT, should be taken into account. These may, for example, be passengers in the participating vehicle or other road-users who are videoed.

Human rights legislation is also relevant, as is the Helsinki Declaration of 1964 and its subsequent revisions. This declaration enshrines the right of the individual to be informed and provide prior consent on a voluntary basis. The individual's protection and rights supersede any interests of scientific progress.

Carrying out an FOT give rise to a considerable number of legal and ethical issues. It is not possible to provide a comprehensive guide to all the legal issues that can arise in a particular FOT, as these may be very dependent on the system(s) to be tested and on the study design adopted. It should be noted that the regulations and laws vary from country to country and that even where there are European laws and regulations the interpretation of these may vary between countries. Thus projects carrying out FOTs in more than one country or carrying out FOTs that potentially involve cross-border traffic may need to consider the legal implications in all relevant countries. Another aspect is that projects fully consider health and safety aspects. It should be noted that not carrying out a prior risk assessment and therefore not giving proper consideration to the safety risks that may result from an FOT can expose not only the participants, but also an organisation to risks.

Legally, when undertaking an FOT, it is very important to understand the need to seek legal advice at national level (i.e. in the respective Member State) taking the concrete test design into account. The first version of the FESTA Handbook already identified this aspect (**Errore. L'origine riferimento non è stata trovata.**) and sensitized further by pointing out a number of legal aspects requiring special attention (from the author's point of view – a German lawyer). It must be made clear that even now – since FOT-Net 2 has additionally compiled legal information from Spain, Italy, Netherlands and France on FOT-relevant issues – the legal advice still remains at an abstract level. The review – which has led to a substantial increase in details on these countries – will only allow for a first estimate of what needs to be considered. On the other hand this information is the basis for planning subsequent

procedures in timelines, estimating legal costs involved, defining measures that can realistically be implemented or in the identification of approving authorities (e.g. for data protection, etc.). It must, however, be pointed out that information provided therein is not a final opinion on the law of these Member States since details on the single case are decisive. Furthermore, important differences in national law make it crucial to involve legal expertise at the level of every Member State affected by the FOT and on the basis of predefined data logging facilities and specific systems/applications where relevant.

In terms of the project timeline, legal and ethical issues need to be considered in parallel from the beginning to the end (and indeed afterwards in terms of data protection, especially if the data are planned to be re-used in new projects, see FOT-Net 2 -D3.2). So the discussion in so far does not neatly follow the FOT chain.

As these issues are complex and it is not always easy to find a good solution, it may be useful to consult (next to legal experts) other, comparable FOTs that dealt successfully with them successfully and learn from their experiences

The FOTIP, discussed in Chapter 2 and presented in table form in the Annex B, provides information about when in the FOT process the various legal and ethical issues need to be considered. The project plan needs to clearly identify who are the persons responsible for ensuring compliance or involving legal expertise in the respective Member State.

3.2 Participant recruitment

In recruitment it is essential to ensure that participants have legal entitlement to drive the vehicles in question and are eligible for insurance. It may be wise to have insurance coverage for the fleet as a whole. If the participants are to drive their own vehicles or vehicles that belong to a fleet not under the control of the handling organisation, then insurance coverage needs to be confirmed. Coverage when travelling to other countries may be relevant.

In some countries, it may be a requirement for the participants to undergo a medical examination to prove their capability to take part. In any case, it would probably be sensible to ascertain if they have any medical conditions that might affect their ability to participate. However, in some countries data privacy requirements will not allow to ask about medical details so that this will then need to be solved by positively informing on requirements for participation in participant agreements as FOT/NDS termination criteria (for the individual participant).

3.3 Participant agreement

There is a need to formalise the arrangement between the organisations responsible for the relationship with the participants and those participants themselves. This arrangement is legally a contract according to civil law. Even when this arrangement is made in the form of e.g. “a letter of agreement”, this will not alter the legal character of the document which influences the legal relationship between the participant and the handling organisation. Issues potentially relevant between the handling organisation and the test participant should be regulated in a contract to provide for legal certainty on both sides (e.g. on

obligations, liabilities, insurance issues, information on the logging of personal data requiring informed consent, which parties will use the data, data sharing after the project including the use of personal data (see FOT-Net 2 - D3.2) etc.). In this respect, legal validity of the arrangements must be considered. This is an issue touching the general rule of freedom of contract as a limiting exception. By way of illustration the common limitation that the handling FOT-organisation cannot prompt participants to agree (in a legally effective way) to fully exclude liability, may in so far serve as an example common in several EU Member States. Again the legal situation in this respect can differ strongly which requires dedicated legal expertise at Member State level to achieve legal certainty. A lawyer can provide advice on this and should definitely be consulted. Further issues of relevance to be considered and covered in this agreement are what shall happen in the event that a participant commits a traffic offence and/or incurs a traffic penalty (speeding ticket, parking ticket, etc.). Another is who is responsible for minor damages to the vehicle and payment of any insurance excess. The issue of who is allowed to drive, e.g. other household members, and under what circumstances also needs to be considered. Only the participants will have been properly informed about their responsibilities. There is no way to ensure that any third parties are properly briefed.

3.4 Data protection and data ownership

Data protection in general is stipulated by an EU directive of 1995 and is enshrined within the national laws of the various member states. These national laws may state specific requirements. There is no doubt that an FOT will give rise to data protection and privacy issues. The general approach to data acquisition in the majority of FOTs will be to ask the participant for his/her informed, voluntary consent. The information provided thereby describes the scope of processing intended over the lifetime of the data , including potential re-use of data after the project. No additional processing (e.g. disclosure of the data that might lead to identification of the persons involved) can normally take place without prior consent.

Another issue of high relevance in respect to video data is the issue of cameras directed to the vehicle surroundings: There are two legal privacy issues associated with such cameras: the recognisability of other road users and registration numbers. Both are definitely critical in respect of data privacy regulations in place in several EU-Member States (Annex A and FOT-Net 2 -D3.2). A solution that will avoid complications is low camera resolution that does not allow for any identification of the abovementioned items. Otherwise a solution must be sought at national level either by contacting the national data protection agencies or (possibly as a first approach) by involving the respective legal advisor for the project at Member State level.

Obviously video data must be treated with special care in the information society: Any video recording can potentially be made available over the internet or passed on to other third parties. This will obviously cause substantial problems. It is therefore important that personnel handling and analysing the data are given appropriate education on personal integrity issues.

Furthermore, with respect to video recording (and also audio recording), passengers will not normally have given prior consent to being recorded, so it is questionable whether it is appropriate to have in-vehicle cameras with coverage of the passenger seats. Special issues arise in case of minors. As far as video data is affected, it can be a possible solution to apply very low camera resolution not allowing for identification of persons. More details are provided in Annex A . Legal advice at Member State level needs to be sought in this respect if applicable.

The data server must be protected from intrusion, and normally any personal ID information should be kept completely separate from the main database and stored with additional protection such as encryption. It has to be recognised that, even when data has been anonymised, it may be possible to deduce who has participated, e.g. from GIS data in the database. All data transfers including personal data should be made using encryption.

Data ownership and data sharing relates to stakeholder interests. We have already mentioned the participant agreement, where the participant can agree to re-use of the collected data. Some stakeholders will regard data as strategic or sensitive. For example data can be used to compare systems, and this is usually not in the interest of the system producers or OEMs while on the contrary for policy-makers and road operators the effectiveness of specific systems is an objective that is relevant. To deal with these stakeholder interests, agreements on how to address these issues should be proposed as far as possible in advance. This can be done in three ways:

- Agreements on how to deal with data ownership and re-use as such.
- Procedures on how to change or introduce new research issues based on the collected data. This includes also procedures for potential re-use of the data by third parties.
- Address ownership of data in the tendering procedures or contracts with the (public) organisation providing the grant.

Data collected from the CAN bus represent a special case. Some of the data may reveal information that is confidential to the manufacturer, who may not want to share these data with third parties. Proper data filtering could be implemented in order to make available to the relevant partners only the data that are necessary to the FOT analysis.

An overview and recommendations addressing the abovementioned issues, also from a data sharing and re-use perspective, can be found in FOT-Net 2 D3.2.

3.5 Risk assessment

The project needs a comprehensive risk assessment plan and will need to be able to demonstrate subsequently that the identified hazards have been properly managed. Organisations will normally have a safety management process for this.

3.6 System safety

It is obviously incumbent on those conducting an FOT to ensure that the equipment that they have installed in a vehicle and the modifications that have been made to the vehicle systems do not give rise to any undue hazards. Hazards can arise from radio and electrical interference (where electro-magnetic compatibility tests should be conducted), from reducing vehicle crashworthiness (installations on the dashboard, interference with airbag deployment, and so on) and from HMI designs that cause distraction. The potential for failures to arise from modifications to and interaction with in-vehicle systems needs to be handled by means of a formal system safety assessment.

3.7 Approval for on-road use

Vehicles are generally subject to Whole Vehicle Type Approval processes and to Construction and Use regulations. Before it is certain that it is legal to operate a modified vehicle on public roads, a check must be made with the appropriate authorities, who may be the national government or a designated approval agency. Once a vehicle is certified to be legal to operate in one European country, it can normally be driven legally in other countries. This again should be subject to the legal advice sought in respect of the specific FOT.

3.8 Insurance

Insurance requirements extend beyond the insurance of the vehicles and possibly of the participants. There is also a need for indemnity insurance to cover the FOT as a whole. This may be provided by an employing organisation's professional indemnity insurance, but it is vital to confirm that the large risks are covered. Insurances and liabilities tend to differ greatly between the EU Member States. It is therefore advisable to seek legal advice in this highly risk-sensitive aspect.

3.9 Responsibilities

There are no very precise rules about responsibilities, but each contributor should be responsible of the component that he has realized or integrated. In the case that an accident occurs, damaging people and/or goods, as normally happens in any such event, an investigation is opened in order to establish:

- The dynamics of the accident (this could be facilitated by the recorded data)
- The cause (driver, third parties, vehicle fault, road equipment fault, road problems, missing signs, ...)
- In the case of driver failure contributing to the accident, the experimental systems may have negatively influenced the driver and then these systems could be indirectly a cause.
- In the case of vehicle fault a complex technical analysis should be made in order to identify the component originating the fault which may depend on design, poor manufacture or incorrect installation.

These issues can be handled safely beforehand on the basis of sound legal treatment (via contract, insurances, etc.). It is therefore important to involve legal expertise at the level of the EU Member States affected by the FOT.

3.10 Video data collection (specific issues)

Video data collection within the vehicle has been covered in section 16. However, there are some additional points to consider. For example, there may be locations encountered where it is illegal or prohibited to video externally — border crossing, military locations, private premises. The possibility of this happening needs to be considered; it is likely to be more of a problem in truck FOTs.

3.11 Ethical approval

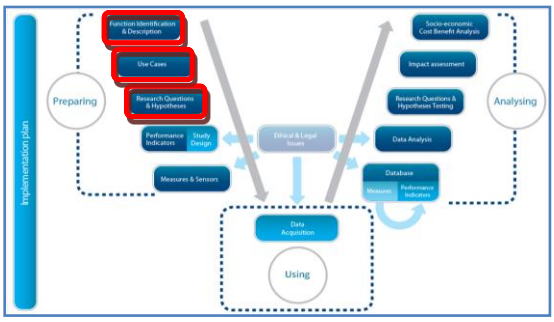
Ethical approval to conduct an FOT may be even more difficult to obtain than legal approval. In many countries and in many organisations there are strict ethical approval and human subject review procedures. However, the information provided so far for Germany (Annex A) and Spain, Italy, Netherlands and France (FOT-Net 2 -D3.2) reveal no strict requirements for an ethical approval.

If ethical approval is required, these procedures can be very time-consuming, so that time for the process needs to be considered in the project plan. Human rights legislation is also relevant, as is the Helsinki Declaration of 1964 and its subsequent revisions. This declaration enshrines the right of the individual to be informed and provide prior consent. The individual's protection and rights supersede any interests of scientific progress.

3.12 Iteration

Considering ethical and legal issues may influence the outcomes of the different phases of the FOT chain. It may be necessary to re-think some phases and to abandon choices made earlier. For example, if it is not possible to collect certain data due to legal or ethical issues, it may no longer be possible to test certain hypotheses or to use certain performance indicators.

4. From Functions to Hypotheses



The final objective of an FOT[FW] is to evaluate in-vehicle functions[FW] based on Information Communication Technology (ICT) in order to address specific research questions[FW]. These research questions can be related to safety, environment, mobility, traffic efficiency, usage, and acceptance. By addressing the research questions, FOTs promise to furnish the major stakeholders (customers, public authorities, OEMs, suppliers, and the scientific community) with valuable information able to improve their policy-making and market strategies. Individuating the most relevant functions and connected hypotheses to successfully address the above-mentioned research questions is one of the major challenges in an FOT. In this Chapter, the process of individuating the functions to be tested in an FOT and

Applying this Process to Naturalistic Driving Studies

This chapter is written from the perspective of FOTs on various kinds of systems. By contrast, NDS investigate driving itself without the constraints of the experimental conditions that are normally required in FOTs. In FOTs, there is a natural progression that starts with a specification of the functions to be evaluated, and then moves on to the environments and situations in which the tests will be conducted and the experimental design that will be employed. The specified research questions follow in a natural progression, and, in forming the hypotheses, there is generally an anticipation of the direction of change, e.g. that a driver assistance system will promote safety or fuel economy. In framing hypotheses for FOTs, we seek to gain a deeper understanding not only of what changes occur but of how they occur, i.e. of how a particular function influences user behaviour.

By contrast NDS are much more exploratory. They aim for deeper understanding of driving behaviour as it relates to safety and in some cases environmental aspects of driving. They can thus be likened to other observational studies in traffic. Of course there is a kind of experimental design even in NDS: the vehicles and participants have to be selected and the choices made here will be determined by the focus of the study and the research topics being addressed.

NDS focus on how drivers manage their driving in changing situations and environments and on how breakdowns in the safe operation of vehicles come about. Hence they are primarily interested in why problems occur and do not occur. This makes research questions, as opposed to hypotheses, the natural focus of NDS. And those research questions will tend to be initially quite broad — what is the relationship between distraction and safety-critical events or why do young drivers have a problem with loss-of-control on curves at night-time. The RQs can then be further specified in the form of sub-RQs and so on. From those sub-RQs, appropriate performance indicators and measures can be defined. This then leads to the specification of the Data Acquisition System and associated subjective and objective data.

One of the challenges in NDS is prioritising the research questions. It is comparatively easy to generate hundreds of RQs and sub-RQs. The difficulty then is likely to be setting the boundaries of the study by determining which ones are high priority and can be addressed within the resources available. However, if the final database is sufficiently broad and flexible, it can become a valuable resource for the investigation of additional topics and RQs.

Those engaged in NDS may benefit from consulting some sections of this chapter, such as the material on situations. Chapter 5 is generally applicable to NDS.

the relevant connected hypotheses will be elucidated. Specifically, the reader will be guided in the process of

1) selecting the functions to be tested,

2) defining the connected use cases to test these functions,

3) identifying the research questions related to these use cases,

4) formulating the hypotheses associated to these research questions, and

5) linking these hypotheses to the correspondent performance indicators[FW].

The FOT chain shows specifically the steps reported above.

The steps may be influenced by other elements of the FOT chain. The selection of functions may be driven by the socio-economic impact that is expected, or by the research questions. When details are filled in later on in the process, it may be necessary to re-visit earlier steps, for example limits in resources or technical capabilities may lead to a decision to limit the amount of functions, use cases[FW] and hypotheses.

4.1 Systems and functions

In the last few years, the number of ICT functions available for use on standard vehicles and more generally while travelling has been rapidly increasing. ICT functions are intrinsically designed to provide the driver or traveller with new, additional information. However, the extent to which this increased amount of information from these ICT functions results in clear and positive effects on safety, environment, mobility, usage, and acceptance in real traffic situation[FW] is unknown. FOTs warrant to evaluate, for the first time, these ICT functions in a real traffic situation during naturalistic driving. In this handbook we refer to 1) in-vehicle, 2) cooperative, and 3) nomadic systems intended as a combination of hardware and software enabling one or more ICT functions. Depending on the different systems implementing a specific function, different challenges may have to be faced during the FOT design.

It is important to note that NDS in the future will increasingly include new systems as they increase in market penetration. For example, future NDS will be able to be used for with-and-without analyses of systems that are new today, but commonplace tomorrow (e.g. Forward collision warning may be as commonplace as ABS tomorrow).

4.1.1 Vehicle systems

Vehicle Systems are a combination of hardware and software enabling one or more functions aimed at increasing driver's safety and comfort. Vehicle Systems promise

1) to increase road safety by increasing driver's attention in potentially hazardous scenarios[FW] (such as the Forward Collision Warning),

2) to improve the driver's comfort by automating some of the operational driving tasks (such as the Adaptive Cruise Control function),

3) to increase driver mobility by furnishing timely traffic information (such as the Dynamic Navigation function), and

4) to increase safety in critical situation by automating the vehicle response (such as the Collision Mitigation function).

Vehicle systems are becoming more and more standard equipment, even in middle class vehicles and commercial vehicles. However, their impact on the driver, the traffic system, the society, and the environment in the short-, but especially in the long-term is not fully understood. FOTs can help quantifying the impact of vehicle systems on driver's workload and to understand how different functions interact with each other in a real complex traffic situation. Further, FOTs will expose these functions to many improbable scenarios which are not possible to be tested during the functions evaluation phase.

4.1.2 Cooperative systems[7]

Cooperative Systems are vehicle systems based on vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I) and infrastructure-to-vehicle (I2V) communication technology. Communication technology has enabled a new class of in-vehicle information and warnings which are more precise in terms of time and location. It is foreseen, that the integration of cooperative systems with autonomous functions will provide a new level of ADAS. Infrastructure based information tells the driver for example what is the appropriate speed to keep on a specific part of road or warns the driver in case of ice on the road or fog.

One of the initial objectives of V2V, V2I, and I2V is to increase road safety. The development of safety-critical V2V systems in Europe has been mainly promoted by the Car-to-Car Communication Consortium (C2C-CC). More recently, the EC has placed emphasis on the application of cooperative systems to achieve environmental and efficiency impacts.

Currently (January 2014), the major stakeholder umbrella organisations are working together to achieve the objective, stated in the MoU signed by the car manufacturers inside C2C-CC, to start real life application of cooperative systems by 2015.

A key point in cooperative systems is standardisation. The minimum set of standards needed for Day1 implementation and requested by the Mandate (N. 453) issued by European Commission are going to be finalised during 2014 by ETSI and CEN. Activities for harmonisation with U.S. and Japanese standards are on going. (see the COMeSAFETY Website for information).

Cooperative systems differ in several areas to conventional areas, which directly influences the FOT planning and operation.

Firstly, there are currently no cooperative systems available in the vehicle market – this implies that all participating vehicles in the FOT have to be equipped manually with

[7] Because cooperative systems create considerable extra complexities, the text covers them in some detail.

prototype systems. This is costly and makes it difficult to find suitable test vehicle fleets. Also, Cooperative systems may offer new functions previously unknown to the driver. Before starting the FOT, drivers have to be educated on the usage of such functions and the general aspects of cooperative systems.

Another difficult issue lies in the essence of cooperation: all V2V functions rely on more than one vehicle being in communication range, and V2I relies on roadside-stations. Before starting the FOT, it is hard to estimate the number of vehicles in one area and how often vehicles will pass a Road-Side Unit (RSU). This *penetration rate* is a crucial factor in designing and evaluating an FOT.

Current estimates from recent V2X FOTs predict a minimal penetration rate of 10% for V2X functions to show a noticeable effect on traffic safety and efficiency. Given the size of the FOT area and the average distribution of vehicles, the number of participating vehicles should usually be considerably higher.

The penetration rate also influences the *Frequency of Events[FW] (FoE)*. In cooperative systems FOTs, the FoE is roughly related to the number of necessary vehicles for a function. While some functions work with only two vehicles (e.g. a slow vehicle warning), other functions require several more (e.g. traffic-jam ahead warning). For certain functions the combined frequency of events might make a naturalistic FOT unfeasible.

A proper assessment of the penetration rate impact can be derived from dedicated simulations. A simulation environment should consider traffic effects, communication models, applications and their respective influence on each other. Such dedicated simulation environments are able to predict the frequency of events for the developed functions and thus support the FOT design and setup process greatly in cooperative systems.

There are several implications derived from this effect to the test setup and execution of cooperative systems FOTs:

- The number of equipped vehicles is considerably higher or the FOT needs to run considerably longer to collect enough data to be representative.
- It might be very difficult or even not possible to run the FOT uncontrolled for some functions(e.g. Emergency electronic brake light,..)FOT. Controlled testing implies the usage of dedicated tools to specify the test cases, the test scenarios and to run the test. They also need to be linked to the vehicles and drivers to control and monitor the running test. (See 6.5 for more details)

It is most likely that the initial market drive will be for V2I-based applications until there is a critical mass of equipped cars on the road. FOTs can assess the technological and business feasibility of cooperative systems and may be necessary to complete the validation of such systems. In fact testing to validate cooperative systems may require very complicated set-up which may not be possible unless in a real-traffic set-up.

One important issue is the installation of infrastructure devices that could be needed. Permission for the testing and team involvement should be planned in advance considering the impact on normal traffic during the installation phase. In the testing phase proper

logistic support is needed in order to allow technical people to operate with debugging equipment in order to tune the systems. Assessment of RSUs and antennas positioning could be rather time-consuming. Protection and maintenance of road-side equipment should be ensured.

A specific complexity of cooperative systems is due to the distributed nature of involved components. It is not only sufficient to log the vehicle state and the function state, but to completely evaluate the functionality of a cooperative system it is also necessary to handle all incoming and outgoing V2X data and the Local Dynamic Map state on the vehicle side.

These data are provided to or received from other nodes of the V2X network (a node is defined as ITS Vehicle Station or ITS Roadside station depending on the location) and directly influence the system functionality. Consequently the scope of logging has to be widened to cover the data provided by ITS Stations located in the relevant geographical areas. Each ITS Vehicle Station needs to be equipped with a logging device. This device or software component has to be able to write a common log destination of data provided by all important internal sources and the network :

- Vehicles Data (CAN Bus)
- Position and Time (if not present on CAN)
- Networking: Incoming and Outgoing Messages (CAM, DENM)
- Local Dynamic Map Status (List of all neighbour stations, List of all events)
- Function Status (to monitor how applications are working)
- Other Facilities (connection to backend-services, etc.)

It is obvious, that all these various sources provide a huge amount of data. This is a major consideration for the logging system on each station and should influence it in two ways:

- Qualitative: Each log entry should be reduced to the minimum (entry type ID, timestamp, value). It is possible to further reduce the required space, e.g. by only writing the time and not the date into each entry. It can also be useful to compress the written files. It is strongly advised to use a “strict” logging scheme, which only allows pre-defined log entry types, as this simplifies post-processing and also allows saving disk space.
- Quantitative: Logging does not have to record all values all the time. Although data storage cost is lowering more and more, if necessary, it is possible to reduce the used space drastically by using pre-defined logging profiles. These are generated with a mapping from those performance indicators needed for evaluation to the available measurements. For each test a logging profile can be pre-selected, which contains only those measurements needed.
  In more complex settings the logging system can also automatically switch profiles depending on the measures values (a critical situation can be detected from measurements and a wider logging profile can be selected for a given time). It is also possible to select the log profile directly in controlled testing to guarantee that all necessary measures are logged. The quantity of data to be collected is discussed in

section 9.2 where two approaches ("space mission" and "minimum data set collection") and related issues, are presented.

The ITS Roadside Station provides similar measures (except vehicle data) and it is advised to use a similar architecture for the logging system. In contrast, the measures available to the ITS Central Station can differ widely and are specific to each FOT. Roadside sensors may be needed to provide information related to interaction with non-connected vehicles and other situational variables.

For controlled testing scenarios it is advisable to include monitoring features in the logging component. Monitoring is a real-time transfer of selected information available to the logger. In a central counterpart to monitoring this information is displayed to visualize the current test progress. Typically the monitoring value set is a subset of logging and should at least include vehicle ID, position, heading and speed. In this minimal set the test operator can follow vehicle progress through the test. More advisable monitoring sets include direct information from developed functions to monitor in real-time, if the test intent (to trigger the developed functions in a specified way) has been met. Test operators can thus determine if a test was successful (and the detailed logs are valuable for evaluation) or if tests have to be repeated right away. Sophisticated test control and monitoring tools support operators in this task.

Log files in cooperative systems FOTs can quickly take up gigabytes of data. A rough estimate for a typical cooperative system is 500Mb per day per vehicle. It is important to design the test and logging system to cope with this data amount. Log files have to be saved locally on vehicle and roadside station, have to be transferred to the storage and have to be integrated into a central database. Throughout this process it has to be assured, that no data can be lost due to a single point of failure.

The transfer of data from vehicles to the central storage might exceed possible bandwidth from 3G cellular connections. Other possible ways include transfer over Wi-Fi and manual transfer by USB Stick or Drive. Log data should be compressed to speed up transfer times. Depending on the vehicle availability (average run-time over day, time in between access to vehicles) one solution has to be chosen and the system has to be specified in a way, that the in-vehicle storage can fit enough data to hold all log-files until extraction.

When large international FOT are planned, cooperative systems may require interoperability tests. The costs, the organisational and legal aspects (prototypes in foreign countries) should not be underestimated when planning these testing sessions.

4.1.3 Nomadic devices

The use of so-called Nomadic Devices for transport and traffic related applications has become increasingly commonplace in the last few years. The first wave of such devices were dedicated Satnavs, also known as Personal Navigation Devices (PNDs). The second wave of functions have been a large variety of "apps" for Smartphones. In addition some PND suppliers have formed alliances with vehicle manufacturers so that their products may be fitted as original equipment. Whether fitted as original equipment or functioning on

Smartphones, the hardware platforms tend to operate autonomously of the vehicle. Indeed many of these devices can provide traveller functionality — for pedestrian route-finding, for public transport information, to locate points of interest — outside the vehicle.

The functionality of PNDs and Smartphones has been evolving over time. Initially, the functionality of PNDs was focussed on guiding the driver from geographic point A to a point B, using buttons and the screen to set the desired destination. That functionality has now evolved so that the devices may offer a media player, media viewer, mobile phones integration, and may be voice activated. They can also provide routing and navigation support outside the vehicle. Smartphones started from an emphasis on telephony, supported by additional functions such as an address book and texting. Additional functions(photography, navigation, radio, media centre, etc.) were then continuously added. Now that GPS is generally provided, they are navigation-capable, so that navigation apps are commonplace. On the other side, PNDs are broadening in capability and now tend to include a mobile phone SIM card. Thus there is a large overlap in capability and functionality.

This provides a capability to download map updates but also gives cooperative-system-like functionality in the form of live traffic updates and other environmental information. The same mobile phone link can be used by service providers to acquire real-time traffic conditions. Smartphones, with their built-in networking capability, can offer the same functionality.

As well as links to mobile networks, nomadic devices are now available with Bluetooth, Wi-Fi networking, assisted GPS, speech recognition and high-end operating systems with the capability to provide dynamic geo-referenced applications to assist the driver and traveller. Specifically tailored aftermarket devices, sometimes offered by the same manufacturers as personal PNDs, are now targeted at the fleet market for management of logistics and other fleet functions. Some of these devices provide links to CAN data. An increasing focus of such systems is fuel economy and feedback to drivers on their efficiency in driving. Another sector for aftermarket devices is the pay-as-you-drive insurance market.

Nomadic devices need to be evaluated from the perspectives of user behaviour and acceptance, safety (particularly in regard to HMI issues), travel and traffic impacts and environmental implications. It also needs to be recognised that such devices can have broad mobility implications, both in terms of the strategic level of driving (route choice) and in terms of trip generation and mode choice. Any evaluation of usage needs to consider the potential for both in-vehicle and out-of-vehicle usage of these devices.

One of the critical characteristics of PNDs and Smartphones is how far the device is integrated within the vehicle. Many devices use specific mounting kits for in-vehicle installation (connection to power supply, GPS and in some cases the vehicle's audio system). Typically PNDs are mounted with a suction cup directly to the windscreen, while cradles for smartphones may also attach via a suction cup. As a result they may impede the driver's field of view increasing the risk of accidents. There is also often a problem with small screen size and inadequate audio volume. Nevertheless, the popularity of these devices has not been affected. To increase usability and reduce negative side effects, an Automotive Bluetooth profile has been developed so that such devices can use the vehicle's built-in HMI.

This can offer improvements both in user input to the device (stalks, buttons, speed recognition, link to hands-free mobile phone, etc.) and in device output (using integrated audio, and screen).

The general ease of use of a device will have a major influence on acceptance and willingness to pay. Here ease of use refers not just to the usability while driving but to the user experience in all aspects of usage — pre-trip, in-trip and post-trip. Post-trip functionality is very relevant to usage in the fleet market and to support for and feedback on eco-driving.

Due to the above-mentioned fast innovation cycle, FOTs studying nomadic devices may require state of the art planning in order to keep up with the introduction of new features and functions. They will also need to consider the surrounding infrastructure since (rather like many cooperative systems) many functions rely on information and support from the outside. Weather forecasts, traffic information, updates on road conditions, dynamic speed limit information and speed advice are all dependent on service providers.

4.1.4 Combinations of functions

There are many FOTs which investigate the impacts of a combination of functions sometimes because systems and functions come in a bundle. One such common bundle is the combination of Adaptive Cruise Control and Forward Collision Warning. Both functions make use of the same sensors, and indeed second-generation ACCs generally implement a warning function to indicate to the driver when the deceleration demanded of the ACC in order to prevent a collision with the preceding vehicle is beyond the function's designed capability. In other cases, an FOT may be investigating a function that resides on a platform which offers many other functions. This is almost invariably the case when studying functions that reside on nomadic device such as a smartphone or SatNav. In some cases a project will create a new function and provide it to users on a standard consumer nomadic device. It is not practical or reasonable to demand of users that they do not use the full functionality of the device and attempts to disable features may well annoy participants.

Therefore in planning the evaluation, it is important to consider how functions may interact with each other and how those interactions might affect user behaviour. This needs to be done at the stage of an FOT when research questions and hypotheses are initially formulated. What needs to be considered is:

- Can the effects of the various functions be disentangled? Note that it may not be possible or feasible to do so, particularly if the functions are closely coupled together.
- Does the experimental design need to be modified to enable both the single effects of each function to be investigated as well as the effect of the functions in combination?

Some systems are now so integrated that it is no longer possible to disentangle them completely. An ACC that does not incorporate FCW is no longer on offer. An ACC + FCW system can be driven with the ACC off but cannot be driven in ACC mode with FCW off.

Indeed disabling the FCW functionality within an ACC could be considered unethical and may be impossible for practical reasons.

Where bundles cannot be disentangled, it is only possible to investigate and report on the impacts of the combined systems. But when, for example, there are separate apps on a nomadic device, then consideration needs to be given to how those apps might interact. The investigation can be carried out in a naturalistic manner — i.e. the participants are free to choose when and when not to use the applications and functions that are not the immediate focus of the FOT. Alternatively, it can be carried out in a more experimental manner by means of instructions to participants, systematically enabling and disabling the functions, or carrying out controlled drives with the functionality of the system(s) carefully manipulated.

In formulating hypotheses, it is useful to think both of the singular effects of a system or function and of the synergistic effects that one function may have on another. Thus the recommended procedure is to start with the individual functions and then to proceed to combinatory effects. The process should therefore be to:

1. Begin separately with the individual functions, generating a list of research questions and hypotheses
2. Examine commonalities and conflicts between the systems and functions and derive hypotheses from those commonalities and conflicts:
    a. Can they generate simultaneous messages and/or warnings?
    b. Do they have a common interface and can they both be activated simultaneously?
    c. Are the same performance indicators relevant to each system and/or function?
    d. Are there common factors influencing usage?
3. Distinguish between hypothesis additive effects when the two systems interact with each other and multiplicative effects when the presence of a second system will alter the effects of the first. Additive (or subtractive) effects mean that the size of the effects will change. Multiplicative effects mean that the relationship is different, i.e. that there is an *interaction* in statistical terms. Consider situations in which such multiplicative effects might be important.

The application of this procedure should produce a comprehensive set of hypotheses on how the functions might interact and should affect the subsequent experimental design. As in all such work on the preparation of research questions and hypotheses, the reasonableness of pursuing every possible combination in a structured experimental design needs to be considered. Any additional function can impose a huge cost in terms of the increase in the number of possible combinations. It can be argued that there is a rationale also for multiple baselines, i.e. with all functions off, and with one function at a time off.

Complex experimental designs have large practical costs associated with them and the benefits of such designs need to be carefully considered. The costs can be in the form of the

number of different baselines that may be required and of the time needed for data collection on each combination. A full experimental design in which ordering effects are considered may well be totally impossible for practical reasons. This can all lead to excessive time required for FOT execution. But another side effect can be the sheer difficulty of getting the participants to comply with all the different conditions of the experimental design. These are arguments for using a more naturalistic approach in which the participants are free to use whatever combination of systems and functions they choose. Of course system and function state will need to be recorded. This naturalistic approach has some disadvantages:

- Not all combinations may occur and not all participants may experience each combination
- It may be hard to take care of seasonal effects
- There may be insufficient sample sizes in some conditions so that experimental power is inadequate

However, the naturalistic approach also has advantages in that:

- Participant compliance will generally be assured
- The frequency with which the various combinations are used can provide useful information for the scaling-up process

Alternatives which offer greater efficiency than a full experimental design can be proposed. Laboratory experiments on driving simulators can be adopted to examine synergistic effects and a priori analysis can be applied at an early stage in an FOT to identify combinations that are of particular interest and which should therefore be the focus of attention. This can lead to a satisfactory but incomplete experimental design.

4.2 General methodology

The main advantage of an FOT is that it has the potential to give insight into system performance in naturalistic driving situations, as free as possible from any artefact resulting from noticeable measurement equipment or observers in the car. Therefore the first step when planning an FOT is to identify systems and functions where considerable knowledge about their impacts and effects in realistic (driving) situations is of major interest, but is still lacking (see Section 4.2.1). Another domain for FOTs is the area of systems and functions which need a certain penetration rate to work at all, like especially cooperative system.

After the identification of the functions and system, which should be tested in an FOT, the goal is to define statistically testable hypotheses and find measurable indicators to test the hypotheses. To reach this goal, several steps need to be taken, starting from a description of the functions down to an adequate level of detail (see Section 4.2.1). This means that the main aspects of the functions, its intended benefits and the intrinsic limitations have to be described to fully understand objectives and limitations and to derive reasonable use cases.

Secondly, these use cases need to be defined (see Section 4.2.2). Use cases are a means to describe the boundary conditions under which a function is intended to be analysed. A

general starting point is given by the functional specifications from the function description part. But it might also be of interest how a function performs when certain preconditions are not met and to identify unintended and unforeseen effects.

Starting from the use cases definitions specific research questions need to be identified (see Section 4.2.3). Research questions are general question to be answered by compiling and testing related specific hypotheses. While research questions are phrased as real questions ending with a question mark, hypotheses are statements which can either be true or false. This will be tested by statistical means (see Chapter 9). One might already have a very clear idea from the beginning which hypotheses are to be tested in a very specific situation during the FOT. However, this very focused view might result in an extreme limited experimental design, where important unintended effects will not be considered. The process to define hypotheses developed in FESTA aims to prevent these potential issues (see Section 4.2.4).

Finally, hypotheses can only be tested by means of reasonable indicators (see Section 4.2.5).

These steps are shown as parts of the complete FOT and are elaborated further in the following sections. FESTA deliverables D3, D4 and D5 provide additional detail on the application of the FESTA methodology to identify functions and systems and to develop hypotheses for the experimental design. All steps, from the description of the systems and functions, the development of use cases and scenarios, as well as the research questions and hypotheses and the proposal of related performance indicators are covered.

In the FESTA methodology, function description is a starting point. However, this may pose some problems. Not all FOTs are testing one pre-defined function; sometimes a set of functions or systems are to be evaluated. Or in naturalistic driving studies, the focus of the study may not be on functions but on driving behaviour in general. Stakeholders may have different ideas about the functions they want to test, and function descriptions are not always clear.

One issue is whether a function is generic or manufacturer-specific. In other words, how far should the particular manufacturer's specification and functionality be considered. The specifics of the function can clearly have an impact on acceptance and behaviour. An example here might be an LDW that gives auditory and visual warnings and a different one that provides haptic feedback through the steering wheel. It is possible that one design will turn out be more effective than the other.

When it is possible to start with a clear function description, this will allow a detailed planning of the data collection plans, of the experimental design and hence of the costs. But it is of course still necessary to further specify research questions, hypotheses and performance indicators. In the case that functions are not well-defined or a decision has not made before the start of the project as to which functions to investigate, it may be more advisable to start with defining the research questions and select functions that seem most useful in answering these questions.

The issue of how many functions can be investigated in an FOT is a matter of the resources to be deployed and of whether the impacts of each function is to be investigated separately

or alternatively whether the functions are to be treated as a package. Multiple functions can have interaction effects with each other and combinations of functions can therefore have impacts which are not simply the sum of the individual effects. On the other hand it may not be feasible to get the functions to work separately — for example FCW is now generally provided in combination with ACC so that when using ACC it is not possible to switch off FCW (see section 4.1.4).

4.2.1 Step 1: Selection and description of functions

Usually it is quite clear from the beginning what functions or at least what type of functions will be the object of an FOT . However, to select the specific functions but also in case the type of functions has not yet been decided, a Stakeholders Analysis is recommended. During this analysis, the needs of the different stakeholders need to be identified and merged into a common requirements description. Stakeholders are those whose interests are affected by the issue or those whose activities strongly affect the issue, those who possess information, resources and expertise needed for strategy formulation and implementation, and those who control relevant implementations or instruments, like customers, public authorities, OEMs, suppliers, and the scientific community. It is of vital importance that all relevant stakeholders are included in the analysis to guarantee that the selection process will not itself bias from the beginning the appraisal of the gained results.

It is recommended to evaluate the stakeholders' needs by means of questionnaires, workshops or well documented interviews of stakeholders' representatives. It is also quite important to describe the selection process sufficiently to prevent from misjudgement.

The basis for all following steps is a sufficient description of the selected functions.

For these purposes it is suggested to collect the necessary information into a spreadsheet, structured in two parts:

- A first one, the functional classification, where a short high level description of the main aspects of the function should be given. This information is usually provided through the system specifications given by the system vendor or OEM..
- The second part of the description should comprise limitations, boundary conditions and additional information which is necessary to understand how the function works.

The boundary conditions part should describe where and under which circumstances the system/ function will operate according to its specifications, where the FOT should take place and which type of data needs to be recorded during the FOT to enable a good interpretation of the results. It consists of:

- Infrastructure requirements, cooperative systems and nomadic devices requirements. Here all required actors besides the actual system need to mentioned, which might have an impact on system performance, service availability or similar. It is intended to trigger the consideration of factors which are external to the system/ function under evaluation;

- Demographical requirements / driver requirements: Especially the user or driver recruitment needs to take into account, whether a function is particularly designed for a specific group of users or drivers. Drivers differ on a large variety of characteristics, which may all have an influence on how they drive and use different systems and services. These differences may be important to take into account when planning an FOT. Four categories of driver characteristics may be distinguished:
    - Demographic characteristics: gender, age, country, educational level, income, socio-cultural background, life and living situation, etc.;
    - Driving experience, and driving situation and motivation: experience in years and in mileage, professional, tourist, with or without passengers and children etc.;
    - Personality traits and physical characteristics: sensation seeking, locus of control, cognitive skills, physical impairments or weaknesses etc.;
    - Attitudes and intentions: attitudes towards safety, environment, technology etc.
- Geographical Requirements/ Road Context: This description is necessary for systems which, concerning their functionality, depend strongly on the horizontal or vertical curves of the road layout or on the road type. For example, certain speed limit information systems depend largely on the availability of speed limit information in a digital map, which is up to now only commercially available on high class roads.
- Geographical Requirements/ environmental restrictions: Certain systems are especially designed for specific environmental conditions or, on the other hand, specifications might indicate that the system under evaluation will not work under certain environmental conditions. In this case the location of the FOT needs to be selected carefully and the relevant data must be recorded during the FOT. For example, most of the functions using perception system will be affected by adverse weather conditions. If this is the case it is necessary to log relevant data and take it into account for later data analysis.
- Geographical Requirements/ Traffic Context: The performance of certain systems might depend on the traffic context, that is, the traffic density (e.g. given by the Level of Service) or might even be designed to work in specific traffic densities only. Like the other geographical requirements, this needs to be taken into account when an FOT is planned, performed and the data is analysed.
- Other Limitations: All other limitations need to be mentioned, which might have considerable impact on the performance of functions or systems, since these limitations have major impact on the experimental design and data analysis.

4.2.2 Step 2: Definition of use cases and situations

FOTs will typically test technically mature ICT systems. Therefore, systems and functions to be tested are on the market or close to market and can be easily implemented. But the list grows too long if all possible implementation variations and technologies are considered separately. The use cases put the systems and functions at a suitable level of abstraction in order to group technology-independent functionalities and answer more holistic research questions described later.

Table 4-1: Use Cases, Situations, Scenarios, and their mutual dependence.

Subject	Definition	Comment	Example
Use Case	A specific event in which a system is expected to behave according to a specified function	A use case is a system and driver state, where "system" includes the road and traffic environment.	Car following
Situation	One specific level or a combination of specific levels of situational variables	Thus a situation is a state of the environment.	Rainy weather + darkness + motorway driving
Scenario	A use case in a specific situation	Use case + situation = scenario	Car following on the motorway in rainy weather and darkness

A use case is a textual presentation or a story about the usage of the system told from an end user's perspective. Jacobson et al (1995) defined use cases as follows: "When a user uses the system, she or he will perform a behaviourally related sequence of transactions in a dialogue with the system. We call such a special sequence a use case." Use cases are technology-independent and the implementation of the system is not described. Use cases provide a tool for people with different background (e.g. software developers and non-technology oriented people) to communicate with each other. Use cases form the basic test case set for the system testing. There are number of different ways to define a use case. Use cases in FESTA are very general descriptions, like e.g. "car following". This general description needs to be refined to a reasonable level of detail. This refinement is done by describing so called situations (see Table 4-1). It is the detailed scenario description which triggers the development of specific hypotheses for later analysis.

The situational descriptors are selected in a way that relevant information can be gathered to distinguish between main differences while evaluating systems. The situational descriptors can be distinguished as static and dynamic, where static describe attributes which will not change significantly during one ride of the vehicle, such as age or gender of the driver. Nevertheless this information needs to be stated, since it is one of the main inputs to filter the huge amounts of data in the later stage of data analysis. The second type of attribute is dynamic, since it can change during a ride of the vehicle, such as the system

action status (system on or off), the traffic conditions, road characteristics or the environmental situation.

The situations are defined as a combination of certain characteristics of a use case. Situations can be derived from use cases compiling a reasonable permutation of the use cases characteristics. The identification of possible situations is covered from three viewpoints:

1. systems and vehicle specification,
2. environmental conditions specification and
3. driver characteristics and status specification.

The situational descriptors in FESTA conform to the following structure:

IDENTIFICATION AND DESCRIPTION

Use case name	A name for identification purposes.
Description	General description of the use cases with necessary depth of information to get a quick overview.
Occurrence	Information about the anticipated quantity of occurrences has implications for the amount of data to analyse.

SYSTEMS AND VEHICLES

System status	Depending on the hypotheses the analysis might concentrate on situations where the system is activated or present. *Example: ON/OFF (baseline) or IDLE/ON/OFF*
System action status	Depending on the hypotheses the analysis might want to compare the driving performance between different system statuses, e.g. whether the system is actively controlling the vehicle or not. *Example: acting/ not acting (meaning e.g. ACC controlling car speed or not)*
System/ function characteristics	Depending on the hypotheses an analysis of system or driver performance with respect to special system/function characteristics might be conducted, e.g. examining differences in system performance between nomadic devices (phone, Smartphone, PND...) or effects that depend on vehicle type. *Example: passenger vehicle/ truck/ bus*
Interaction between systems	system and especially driver behaviour might change depending on whether the system under evaluation is the only active support system or whether interactions between two or more systems are foreseen. *Example: interaction between Blind Spot Warning and Lane Departure Warning.*

ENVIRONMENTAL CONDITIONS

Traffic conditions	Performance of some systems might differ depending on traffic density. Others might only be reasonable with a minimal traffic density.

	Example: Level of Service A and B [8]
Environmental situation	System performance differs depending on lighting and weather conditions like rain/ snowfall/ icy roads, etc. *Example: normal/ adverse weather conditions*
Road characteristics	E.g. type of road gradient, super elevation, curvature, curviness, since some systems are dedicated to improve driving performance in curves etc. *Example: urban roads/ rural roads/ highways*
Geographical characteristics	Information about geographical characteristics relevant for the systems. *Example: mountainous/ flat areas, metropolises with high street canyons.*

DRIVER CHARACTERISTICS AND STATUS

Driver specification	Characteristics of the users have an impact on the driving performance. Even if no specific impacts are expected of certain characteristics, some outcomes may be explained better with more knowledge about the participants. A minimum set of data such as age, gender, income group and educational level is easy to gather from participants. Information about driving experience is also important. In section 5.5.6 more details are reported about driver characteristics and about usage of questionnaires for better understanding of driver behaviour and acceptance evaluation.
Driver status	Mindset of the driver *Example: attentive/ distracted/ impaired*
Purpose, distance, duration	Describes the different attributes of a trip (time between ignition on and ignition off). All three aspects have an impact on driver behaviour and hence on patterns in the data.

A set of basic rules has been set for the design of the situations for an FOT:

1. Complementary: situations are not allowed to overlap.
2. Entirety: the sum of all situations should describe the complete use case.

[8] Typical classification used by Highway operators: see e.g. http://www.virginiadot.org/projects/resources/5-LOS_descriptions.pdf

3. Baseline: The same situation without the use of the systems (system off or non-present) is defined as the baseline. The baseline is the basis for the benefit assessment of the system and the comparison between systems. Therefore, for the same use case, there can be many baselines depending on the number of situations.
4. Comparability: functions compared in an FOT need to have the same use cases and therefore same baseline and situations.
5. Variability of situation parameters: depending on the point of view (user, trip, vehicle, single FOT, multiple FOTs, etc…), attributes describing a situation can vary considerably or not.

This list is non-exhaustive and might be extended if necessary.

Finally, out of all the possible situations, one will need to select the relevant ones for scenarios of interest in an FOT. The scenarios are defined as a use cases in a specific situation and therefore one or more scenarios should be considered from each use case. All other situations should be considered out of the scope of the FOT study. However, if possible data should still be collected in all situations in case an alternative study would like to reuse the same data.

The process of defining the use cases will help the FOT for the next steps: the definition of the research questions and hypotheses and finally the identification of the needed indicators. The scenarios as they are defined at this stage of the FOT are not detailed enough for data analysis purposes. For this reason, after the definition of the indicators, the scenarios (and their situations) will need to be further described in terms of events for data analysis purposes. Only then, the scenarios can be classified with a quantitative measurement tools in function of the defined indicators.

4.2.3 Step 3: Identification of the research questions

The research questions specific to an FOT can only be identified once the overall goal of an FOT has been established.

In general terms the goal of any FOT is to investigate the impacts of mature ICT technologies in real use. The core research questions should therefore focus on impacts but there are other questions that 'surround' this core. The range of possible questions is listed below. This list below should be considered a first step in any FOT and not a comprehensive set of questions.

LEVEL OF SYSTEM USAGE

Which factors affect usage of the functions? Examples are

- Purpose of journeys where system is used
- Types of road on which system is used
- Traffic density
- Headway
- Weather condition
- Ambient lighting

How do driver characteristics affect usage of the functions? Examples are

- Personal characteristics (e. g. age, vision)
- Socio-economic characteristics (e. g. family, friends, employment status)
- Journey-related characteristics (e.g. other car occupants, shared driving)

IMPACTS OF SYSTEM USAGE

What are the impacts on safety?

- Exposure
- Risk of accident or injury
- Incidents and near accidents[9]
- Accidents

What are the impacts on personal mobility?

- Individual driving behaviour
- Travel behaviour
- Comfort

What are the impacts on traffic efficiency?

- Traffic flow (speed, travel time, punctuality)
- Traffic volume
- Accessibility

What are the impacts on the environment?

- CO_2 emissions
- Pollutions
- Noise

IMPLICATIONS OF MEASURED IMPACTS

What are the implications for policy?

- Policy decisions
- Laws, directives and enforcement
- Future funding
- Public authority implications
- Emergency service implications

What are the implications for business models?

- Predictions for system uptake
- User expectations
- Pricing models

What are the implications for system design and development?

- HMI design and usability
- Perceived value of service
- Device design

[9] See definition of Accident and Incident in the notes reported in section 1.1.2

- Communications networks
- Interoperability issues

What are the implications for the public

- Public information/education
- Changes in legislation
- Inclusive access to systems
- Data protection

When defining the research questions the trade-off in terms of process and consequences between an in-depth analysis of few research questions vs. general analysis of many research questions should be considered (especially for uncontrolled NDS data), bearing in mind that focus on few research questions often produces scientifically sound results.

4.2.4 Step 4: Creation of hypotheses

Once the key research questions for the FOT have been identified, hypotheses can be derived. The process of formulating hypotheses translates the general research questions into more specific and statistically testable hypotheses.

FESTA distinguishes between more general and open research questions and more specific hypotheses. The definition of a research questions in the case of an FOT is "a general question to be answered by compiling and testing related specific hypotheses". An example would be: "Does having a Forward Collision Warning system improve safety in driving?"

A hypothesis is here defined[10] as:

> "a specific statement linking a cause to an effect and based on a mechanism linking the two. It is applied to one or more functions and can be tested with statistical means by analysing specific performance indicators in specific scenarios. A hypothesis is expected to predict the direction of the expected change."

The term "function" is used because a particular system may have a number of distinct functions — for example, one system could provide both Adaptive Cruise Control and a Forward Collision Warning. It is also the case that one function can be provided by different systems. An example of a hypothesis might be: "Forward Collision Warning will have the direct effect of an increase in minimum Time to Collision (TTC)."

There is no process that can assure that all the "correct" hypotheses are formulated. To a large extent, creating hypotheses is an intuitive process, in which a combination of knowledge and judgement is applied. Nevertheless, a number of recommendations can be made about how this process should be conducted. These recommendations have been

[10] The definitions used here and below are from the glossary of the EuroFOT project.

tested in FESTA and FOT-Net workshop and modified based on the experience of and feedback from FOTs.

Two complementary ways to develop hypotheses have been used. Both ways need to be followed, while it is not of importance which step is taken first. One of the steps follows the sequential check of specific areas in which functions can have an impact; the other step is fully based on the description of specific scenarios. While the one step results mainly in general hypotheses, the other step triggers the development of very specific hypotheses in specific driving situations or scenarios.

4.2.4.1 Top-down approaches

The 6 areas approach

The six areas of impact defined by FESTA are based on Draskóczy et al. (1998). Although this approach was originally designed for formulating hypotheses on traffic safety impacts, it is in fact equally applicable for efficiency and environmental impacts.

The six areas are:

1. Direct effects of a system on the user and driving
2. Indirect (behavioural adaptation) effects of the system on the user
3. Indirect (behavioural adaptation) effects of the system on the non-user (imitating effect)
4. Modification of interaction between users and non-users (including vulnerable road users)
5. Modifying accident consequences (e.g. by improving rescue, etc. — note that this can effect efficiency and environment as well as safety)
6. Effects of combination with other systems

It is not of particular importance to which of these areas a particular hypotheses *is allocated.* The six areas are instead to be used as a checklist to ensure consideration of multiple aspects of system impact.

In applying this procedure, it should be noted that:

- Area 1 includes the human-machine interaction aspects of system use
- The **driving task** (see Figure 4-1) can be defined, following Michon (1985) into the three levels of strategic, tactical (manoeuvring) and control aspects. All three levels need to be considered. All are affected by the input from external conditions, which because they are external to the driver can include the vehicle and devices within the vehicle.

  The *Strategic Level* covers longer-term planning and responses to traffic conditions includes potential modifications to:
  - Mode choice

- Route choice
- Exposure (frequency and/or length of travel)

The *Tactical (manoeuvring) Level* includes potential modifications to speed choice and the effects of such modification on manoeuvring and interaction with other road users. These typically happen in time spans of seconds.

The *Control Level* also includes potential modifications to speed choice and the effects of such modification on vehicle control. Such modifications can occur in milliseconds.

- Consideration should be given to such **mediating factors** as user/driver state, experience, journey purpose, etc.

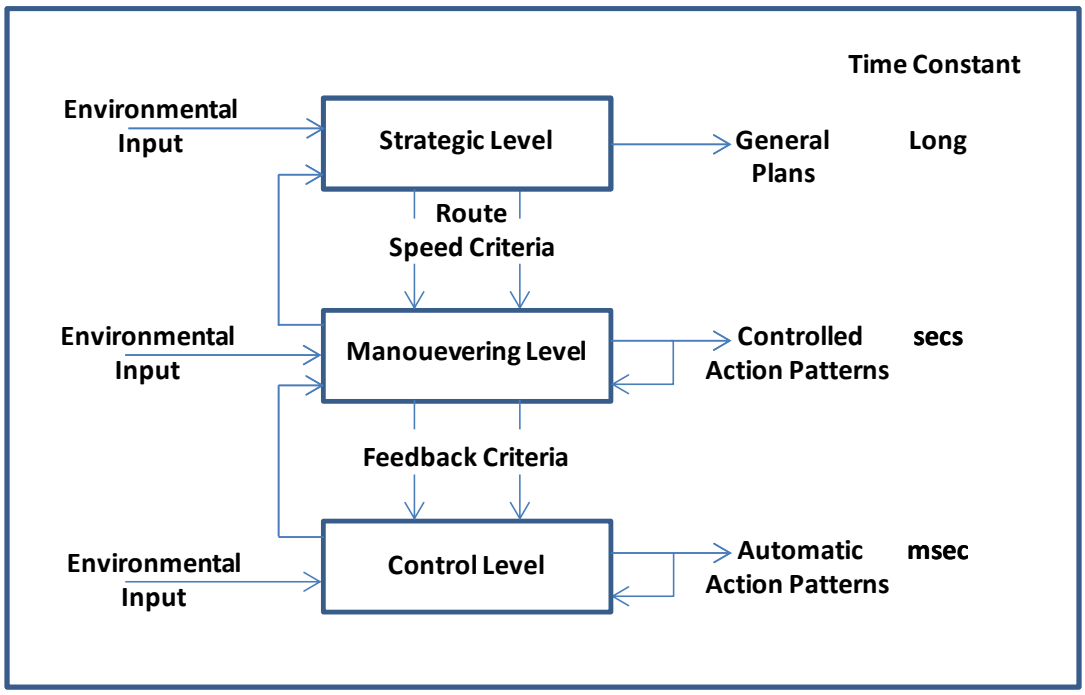


Figure 4-1: The three-level model of the driving task, based on Michon (1985)

It should also be noted that the **effects** of system use may be:

- Short-term or long-term in terms of *duration*
- Intended or unintended in terms of *system design*

Impact area approach

Another useful top-down approach starts from the most relevant impacts areas which are:

Efficiency, Environment, Mobility, Safety and User Uptake.

The basic principle for generating hypotheses using this top-down approach lies in a theoretical understanding of the factors that influence the different impact areas. It should be noted that there is likely to be overlaps of these factors among the impact areas under consideration and hence the same research questions and resulting hypotheses will be

applicable across more than one impact area. The approach will result in *generic* research questions that are independent of the any system functionality.

The procedure for generating hypotheses in this top-down approach is as follows;

- The impact area should be considered in its entire context and primary measures affecting that area identified.
- Secondary factors of these measures are then identified that can be used to explain the variations in the primary measures.
- Finally the variables affecting the secondary measures are identified.
- The variables identified form the basis of the generic research questions "*Is there a change in the variable?*" and the hypothesis based upon an anticipated effect of the variable "*The variable will increase/decrease.*"

This procedure should be undertaken for each of the impact areas.

Using the Safety Impact Area as an example:

- The primary measures affecting safety would be the "Number of events (accidents, near misses) that occur" and the "Severity of the event".
- Secondary factors affecting the first of these measures would, for example be
    - 'Exposure of the vehicle on the road', 'The driving style of the driver', 'The distraction of the driver from the driving task' and 'Any interaction with the fitted device'.
    - Considering the factor 'Exposure', this can be measured with the following variables: 'Length of journey', 'Number of trips undertaken' and 'Road type used'.

These variable lead to the following research questions

- ✓ *Does the system affect the length (miles) of journeys?*
- ✓ *Does the system affect the duration (hours) of journeys?*
- ✓ *Does the system affect the number of journeys undertaken?*
- ✓ *Does the system affect the road type used?*

This leads to the generic hypotheses that can be tested in a statistical manner. The direction each hypothesis should take (e.g. increase or decrease) is based upon the anticipated effect once the top-down approach is integrated with the bottom-up (system defined) approach.

- ✓ *Journey lengths will increase/decrease when the system is used compared to when it is not used.*
- ✓ *Journey duration will increase/decrease when the system is used compared to when it is not used.*
- ✓ *The number of journeys will increase/decrease when the system is used compared to when it is not used.*
- ✓ *The use of rural roads/motorways/major roads will increase/decrease when the system is used compared to when it is not used.*

4.2.4.2 Bottom-up: the use case approach

This process leads to the development of hypotheses concerning specific scenarios. These scenarios are derived from the combination of Use Cases and Situations (see **Errore. L'origine riferimento non è stata trovata.**). Scenarios should be covered systematically. It is recommended that a structured approach be used in scenario development and that an Excel spreadsheet is used as a record.

4.2.4.3 Prioritising the hypotheses

A complete list of the hypotheses that have been developed should be recorded. If it is considered that some are too trivial or too expensive to address in the subsequent study design and data collection, the reasons for not covering them should be recorded. It should also be noted that there are standardised techniques for observing driving behaviour with manual observers which may be less resource intensive than using dedicated data recording. Observations using such techniques can be carried out at various times during the study, preferably along a fixed route.

A huge number of research questions and associated hypotheses from the top-down and the bottom-up approaches will be developed. A key task is to integrate both sets of hypotheses in the context of each FOT. It is envisaged that the bottom-up approach will form the basis of the hypotheses list for an FOT and that the top-down approach will be used in order to check that nothing significant for a particular impact area has been omitted.

After the integration has taken place, the list of hypotheses is still likely to be large. In order to derive a final, manageable set of research questions and hypotheses that can be applied throughout the various test sites, a cost–benefit approach is proposed. Using this approach, an assessment is made regarding the likely "costs" of collecting the data.

Costs can be represented in terms of effort required to derive a performance indicator expressed predominantly in terms of resources. This should be offset against the likely "benefit" that proving/disproving the hypotheses will have. This is measured by way of the likely contribution towards providing a significant answer the research question and thus the level of contribution to the impact assessment. To some degree, this will depend upon the stakeholder needs and requirements, and therefore a prioritization of their needs should be considered.

4.2.4.4 Summary

The basic set of recommendations is:

1. A structured approach should be applied linking a **top-down** approach at the global system level with a **bottom-up approach** which looks more at system states and what can arise from them. FESTA considers it mandatory to combine the two approaches.

2. A multidisciplinary team should *jointly* develop the hypotheses. A workshop at which participants can brainstorm and debate is recommended to achieve this. Participants in the process should include design engineers, traffic engineers and

behavioural scientists, ideally including both behavioural psychologists and human factors experts.

3. The process should iterate between the top-down and bottom-up approaches. It is not particularly important which is performed first, but it is important to cross-check one approach by using the other.

4. An important output of the process is the initial selection of the performance indicators to be used in testing the hypotheses.

4.2.5 Step 5: Link hypotheses with indicators for quantitative analyses

Some of the hypotheses will already incorporate an indicator which needs to be measured, e.g. a very concrete hypothesis like "The function will increase time-to-collision (TTC)". In this case it is obvious which indicator to choose, while the method to measure TTC might include complicated procedures and/ or costly measurement equipment. Chapter 5 gives an overview about many reasonable indicators. One should consider these indicators when planning the experimental design, since a detailed description how to calculate the indicators from measurements is also provided.

Other hypotheses might be rather unspecific, but still reasonable after rephrasing into testable ones. This rephrasing goes hand in hand with the identification of related reasonable indicators. For example, a hypothesis like "The function will increase lane changing performance" is not directly testable, since "lane change performance" is not an indicator itself. Hence, surrogate measures must be identified to evaluate lane change performance. These surrogate measures or indicators can e.g. be found in publications of corresponding research projects. If appropriate information cannot be found or is not accessible, new performance indicators need to be developed. Those indicators and the measurement methodology must be valid, reliable and sensitive, that is, the measurement must actually measure what it is supposed to measure, they must be reproducible and the measures must be sensitive to changes of the variable. A sensitivity analysis should be performed beforehand during a pilot study to make sure that the new performance indicator is suitable. When one or more surrogate measures have been identified, the initial hypothesis can be reformulated into one or more testable hypotheses. In the above mentioned example, reasonable indicators associated to "lane change performance" might be: use of turning indicator or the number of lane change warnings. The initial hypothesis will then be reformulated into: "The system will increase the use of the turning indicator." and "During the system use, the number of lane departure warnings will decrease.". The next step is then to evaluate how the indicators "use of turning indicator" and "lane departure warnings" can be measured. In this context, Chapter 5 provides useful information.

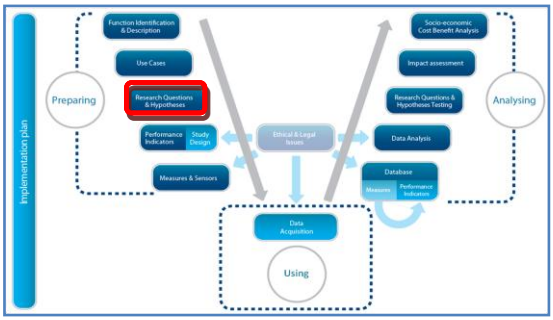
4.2.6 Iteration

Iteration is especially important when defining research questions and hypotheses, because usually a selection has to be made from the large amount of possible hypotheses, both based on their relation with the main impact areas and research questions and on practical issues. Another important iteration point is the impact areas. The final question on the

impact assessment may drive the design of the FOT in all its aspects. When practical issues, such as for example which data-loggers to use, make certain choices hard to realise, iteration to earlier stages is necessary. Cost-benefits analyses and feasibility assessment of different options for the FOT may also drive the design. It is important that there is a good communication between the project members who are in charge of defining research questions and hypotheses and the ones who will be analysing the data in order to assure that the questions can indeed be answered.

5. Performance Indicators

5.1 Introduction



During the process of developing hypotheses[FW], it is important to choose appropriate performance indicators[FW] (PIs) that will allow answering the hypotheses, but that will also be obtainable within the budget and other limitations of the project. Many different kinds of performance indicators have been used in previous studies, and they are related to various aspects of driving. Below a definition and description of performance indicator is given. Further, it is explained how the performance indicator is related to measures, and the types of different types of measures that have been identified are described. Also, examples are provided to illustrate the concepts. An overview of the PI-Measures-Sensors-table, that can be found in the annex of FESTA Deliverable 2.1, is given, along with a background text related to the different groups of performance indicators and measures is provided. Once performance indicator and measures have been defined and linked, it is necessary to test how well they work, which in practice means testing the whole data transmission chain from sensors/devices, vehicles and/or roadside equipment to processed and uploaded data in the research database. The moment to run these tests is the so called "piloting phase", that will be further described in Chapter 6 which is dedicated to Experimental procedures.

5.2 Performance indicators definition

Definition: Performance indicators are quantitative or qualitative indicators, derived from one or several measures, agreed on beforehand, expressed as a percentage, index, rate or other value, which is monitored at regular or irregular intervals and can be compared to one or more criteria.

Further explanations:

- Hypotheses steer the selection of performance indicators and the criteria against which those should be compared. Hypotheses are seen as questions that can be answered with the help of measurable performance indicators.
- Criteria can be baseline, different experimental conditions, absolute values, etc. This depends on the research questions[FW] and hypotheses.
- New performance indicators or combinations can be developed during the course of the study. They will have to be validated in follow-up studies.
- A denominator is necessary for a performance indicator. A denominator makes a measure comparable (per time interval/per distance/in a certain location/…). Therefore "crash" or "near-crash" in themselves should rather be considered as "events[FW]", since they only become comparable when they get a denominator, like "number of crashes per year per 100.000 inhabitants". For certain performance

indicators either time or distance can be used in the denominator (e.g. number of overtaking manoeuvres, percentage of time exceeding the posted speed limit).

For performance indicators measured via rating scales and questionnaires, focus groups, interviews, etc., the "denominator" would be the time and circumstances of administrating the measuring instruments, for example before the test, after having experienced the system, and so on.

Performance indicators are very diverse in nature. There are global performance indicators as well as detailed performance indicators there are observed and self-reported (subjective) performance indicators, there are performance indicators calculated from continuous and from discrete data, and so on. An example of a rather global performance indicators based on continuous log data would be the mean speed on motorways, whereas an example for a performance indicators based on discrete, self-reported data would be the level of perceived usability of a function. Some performance indicators can be based on either self-reported, discrete measures or on logged data, such as for example the rate of use of a system. The participants can be asked how often they use a function, but the actual function activation and the different settings chosen by the driver can also be logged from the system.

All performance indicators are based on measures, which are combined and/or aggregated in certain ways, and which are normalised in order to allow comparisons. The measures are described below.

5.3 Measures

Five different types of measures are identified, namely Direct Measures, Indirect Measures, Events, Self-Reported Measures, and Situational Variables, which are described in more detail below. A measure does not have a "denominator". Therefore it is not in itself comparable to other instances of the same measure or to external criteria. The measure itself, however, can very well be a fraction (like speed). Several performance indicators can use the same measures as input, and the same measures can be derived from different types of sensors. An example would be speed that can be read from the CAN bus, logged from a GPS receiver, or calculated by an external sensor registering wheel rotations.

5.3.1 Direct (raw) measures

A Direct Measure is logged directly from a sensor, without any processing before saving the data to the log file (note that linear transformations like the conversion from m/s to km/h are not considered to be processing). How the sensor arrives at its output is not relevant for the classification. Longitudinal acceleration, for example, is a Direct Measure if logged directly from an accelerometer, but not if derived from the speed and time log. In this case it would be a Derived or Pre-Processed Measure, because it is not directly available from a sensor and has to be calculated from other measures, i.e. pre-processed, before logging. Further examples of Direct Measures are: raw eye movement data, the distance to the lead vehicle as measured by radar and a video film of the forward scene.

5.3.2 Derived (pre-processed) measures

A Pre-Processed Measure is not directly logged from a sensor, but either a variable that has been filtered, for example, or which is a combination of two or several Direct or other Pre-Processed Measures. An example of a Pre-Processed Measure is time to collision (TTC), which is based on the distance between a vehicle and another vehicle or object, divided by their speed difference. The distance to vehicle or object on collision course is a Direct Measure from radar, for example. The speed difference between the own vehicle and the other vehicle or object is another Pre-Processed Measure, based on the own speed as read from the CAN bus, for example, and the calculated speed of the other vehicle or object. Another example of Pre-Processed Measure based on raw eye movement data and vehicle geometry is: pre-defined zones that the driver looks at, e.g. mirror, windscreen and radio.

One major issue with derived measures is that there is to date rather little standardisation on how to calculate them. That means that derived measures calculated in one study may not match those calculated in another study. It also introduces the potential that the calculation procedure adopted may not conform to the best scientific standards and that, as a result, the findings may lack validity. This problem is currently being addressed by the Safety and Human Factors Standards Steering Committee of SAE International, which has adopted a task (Task J2944) termed "Driving Performance Definitions" (http://standards.sae.org/wip/j2944/). The task is focussed on measures related to longitudinal and lateral vehicle control.

A special case of Derived Measures are those that are coded by a human observer after data logging is completed. Examples might be gaze direction coding, classifications of scenarios[FW] or classifications of secondary task engagements. These Measures are considered to be "derived", because data reduction by a human observer is more than only a linear transformation, and they can be based on more than one Direct Measure. In case of secondary task classification one might use both a video of the driver's hands and a log file of an eye tracker, and for scenario classification both a road database and a video of the forward view might be used.

5.3.3 Events

Events can be seen as singularities based on Direct Measures and/or Derived Measures or on a combination of those. They can be very short in time, like a crash, or extended over a somewhat longer period of time, like an overtaking manoeuvre. One or more preconditions must be fulfilled for an event to be classified as such, that is, one or several "trigger" criteria must be exceeded. For the event "overtaking manoeuvre", for example, the non-technical definition might be: A vehicle in a vehicle-following situation[FW] changes lanes, accelerates and passes the vehicle in front, then changes lanes back into the original lane, in front of the vehicle(s) that have been overtaken. Depending on the infrastructure design, the definition might need to be extended to motorways with more than two lanes in each direction, for example.

Several performance indicators can be related to one event type, for example for overtaking manoeuvre it could be of interest to determine the number of overtakings, the duration of

overtaking, the distance/time spent in opposite lane, and so on. For a more technical definition that sets the trigger criteria of when exactly an overtaking manoeuvre starts and when it ends, either the literature has to be consulted or an own definition has to be developed. This can possibly be based on previous data, or, if nothing else is available, on the data from the current FOT.

Events are very important to NDS/FOT studies, because a core type of analysis performed in almost every NDS/FOT is what can be called *Event Based Analysis* (EBA). The basic principle of EBA is analysis is to identify shorter driving segments called *Crash Relevant Events* (CREs, typically in the order of 5-10 seconds), during which the crash risk is judged to be higher compared to other driving, and then to analyse why these events occur, and/or whether their frequency or dynamics change when particular safety systems are made available to the driver.

A key element to NDS/FOT success is therefore defining CREs in a proper way. If the selected events are indeed crash relevant, then extrapolation to the general driver population is both possible and credible. If not, well….

While simple in theory, identifying CREs in NDS/FOT data is a bit more difficult in practice. To begin with, the simplest measure of crash risk, i.e. actual crashes, are incredibly rare, so the final database will usually contain fewer than needed for statistical analysis.

Surrogate events therefore have to be used. These surrogate events need to be driving situations where there is no actual crash, but where the event still unfolds in such a way that its presence can be used as an indicator of crash risk.

A key challenge for all NDS/FOT studies is therefore how to couple surrogate (non-crash) events to crash causation mechanisms. Ideally, one would only use CREs that are known with certainty to be predictive of actual crash involvement. Unfortunately, while a lot of effort has gone into developing algorithms, filtering techniques etc. that allow for efficient yet relevant CRE selection, precise definitions with a clear cut, undisputable connection to crash involvement have yet to be fully established. Selecting suitable CRE definitions is therefore a major decision point for a project, and should be treated as such.

Currently, there are four main approaches to CRE definition in use. These will be given short introductions below. When defining CRE for a project, it is further recommended to read Annex C, which contains a more detailed overview of the pros and cons of each approach, as well as some associated topics to think about.

5.3.3.1 Driver response based CRE definitions

The first approach can be called the “driver response based” approach, and it builds on the general idea that CREs can be identified from the way drivers respond to them. The most common version of this is to look for extreme vehicle kinematics. The basic assumption is that drivers prefer to travel in comfort and generally will not expose themselves to drastic events unless necessary. Thus abrupt velocity and direction changes in the vehicle may indicate unplanned and urgent responses to unexpected situations.

A less common version is to use what can be called the startle response in the driver to find events. The general idea is that unexpected traffic situations that include a perceivable

threat to the driver triggers a response in the form of a general tensioning of the body. This "jerk" may be used as a tell-tale that the driver did not expect the situation and that s/he perceives it as genuinely threatening.

In the driver response based approach, CREs are identified based on how each driver evaluates the situation. For example, while one driver may brake hard at a certain time-to-collision threshold, another driver might not brake at all when at that time-to-collision value. Hence a selection of CREs based on this approach will include the first event but not the second (since the driver did not respond, it is by definition not an event). Driver response based CRE selections will therefore reflect the normal variability in any driver population in terms of driving style, risk perception and capacity to respond. It follows that representative selection of drivers becomes a key issue when using a driver response based CRE selection.

5.3.3.2 Safety function response based CRE definitions

When the study is an FOT, i.e. designed to assess the impact of one or more active safety functions, then a very natural approach to CRE identification is to use the function itself to detect CREs. After all, that is what the function is designed to do. For example, if an FOT is set up to assess the effects of Forward Collision Warning (FCW) on crash risk, the warnings issued can be used as event identifiers.

The downside of this approach is that any CRE that occurs outside the function's detection capacity will be missing from the analysis. It thus becomes impossible to estimate the frequency of CREs which the function in principle needs to detect but in practice cannot. On the upside, a very realistic assessment of function availability and usage is obtained. Since the function only can do something when it is turned on, true availability and usage rates are automatically represented in the data set.

5.3.3.3 Driving context based CRE definitions

A third approach to CRE identification is to base in on driving context. The underlying assumption here is that too small margins equal elevated crash risk. In other words, there exist situations where the safety margins are inherently so small that the slightest mistake or variation could lead to a crash. Prevention of whatever leads to these small margins thus will enhance traffic safety.

The definition of what constitutes too small margins can be static, such as when lane markers are used to indicate boundaries that should not be unintentionally crossed. The definition can also be dynamic. For example, if a vehicle is closer than X to another vehicle and simultaneously closing in faster than Y, it might be considered as being in conflict regardless of whether an action is taken or not.

In a driving context approach, the CRE definition is independent of how drivers resolve the situation. This means all drivers are equally covered, independently of their capacity or willingness to respond. It also means that drivers with a more aggressive driving style will contribute more events per driver to the analysis than those who are less aggressive, since they end up in small margin situations more often. If the analysis team believe that small margins in and of themselves are predictive of crash involvement, this is ok, since these drivers then should have a higher crash risk. If the team is dubious about this assumption however, this approach might not be the right one for the study.

5.3.3.4 Driving history based CRE definitions

A fourth approach is to look for unusual events in a driving history perspective. The underlying assumption is that unusual events in a person's or group's driving history are unusual precisely because drivers try to avoid them. Presumably at least a certain portion of them would be crash risk related.

The advantage of this approach is that it will find the most unusual events that occurred during the study for each person (or group), and it is reasonable to assume that those are events which the drivers would prefer to avoid in the future. The corresponding disadvantage is of course that those events may be special for other reasons than being safety-critical. Even if drivers try to avoid them, they may have little or no connection to traffic safety.

5.3.3.5 Combined approaches

Sometimes projects makes combined use of the approaches above. For example, in euroFOT, after multiple versions had been tried, the CRE definition for lead vehicle conflicts that finally was settled on included the following criteria:

- FCW warning issued (Function based)
- Brakes applied within 5 seconds after FCW warning (Driver response based)
- Max Brake Jerk > 10 (Driver response based)
- Max Brake Pressure > 20 bar (Driver response based)
- Lead vehicle = moving (Driving condition based)
- Direction indicators not in use prior to warning (Driving condition based)

5.3.3.6 Coupling CREs and crash risk – the CRE causation and applicability question

The discussion above illustrate that performance indicators can be built by counting events, or by considering certain aspects of those events. However, it also illustrates that each approach to CRE definition outlined above also represent a different view of how the coupling between CREs and crash risk should be made, and it is not obvious which is the best way to go. What is clear however is that different CRE definitions will lead to different results. In Fitch, Rakha et al. (2008), it is described how another project (Hanowski, Blanco et al. 2008) approached the same data set as Fitch and his colleagues were analyzing with a different CRE selection method. Interestingly, while both projects found hundreds of what they judged to be relevant CREs in that data set, only 7 of the 596 CREs found by Hanowski et al overlapped with those identified by Fitch et al. Clearly, the CRE defined in a project matters.

Thus, FESTA neither can nor will try to provide specific trigger values for Events, nor will the exact measures that have to be included for the definition of a certain event be provided, and the Events listed in the matrix should be seen as examples. Projects must make an informed, explicit, conscious and preferably well documented decision on their own as for

how to use the available approaches to best fulfil their goals, and then set up the CRE detection correspondingly.

Some general recommendations can be made, though. First off, it is important to point out that *not everything can be seen* in the data typically collected in NDS/FOT studies. For example, mental states do not show in video and CAN data, so if the objective is to analyse how intentions, expectations or levels of attentiveness contribute to crashes, probably the study to be conducted should not be a pure NDS/FOT. A mixed approach, where the NDS/FOT is complemented by additional data collection better suited for capturing drivers' intentions and routes, expectations is probably more suitable.

Second, please note that CREs do not have to look like crashes to be relevant for analysis. This comes back to the underlying assumptions about what mechanisms are predictive of crash involvement. If for example it is assumed that high levels of variability in normal driving in and of itself is an accurate crash predictor (a version of the driving context based CRE selection approach above), then each event that captures the tails of the driving parameter distributions is crash relevant, even if it does not look spectacularly dangerous on video.

Third, the selection criteria have to match the time scale of the event to be analysed. This applies in particular to using physiological driver parameters for CRE selection. Many such parameters, like respiratory rate, simply change too slowly for precise correlations with traffic situation changes (i.e. capturing hard brakings by looking at respiratory rate is not likely to succeed).

Fourth, it may strengthen the credibility of the results if the CRE analysis is restricted to injury related CREs. Doing this typically includes finding out at which travel speeds and /or road types injuries occur for the crash type to be analysed, and then limiting the CRE analysis to events which fall inside those speeds and road types.

Again, it is recommended to read Annex C, which contains a more detailed overview of the pros and cons of each approach, as well as some associated topics to think about.

5.3.4 Self-reported measures

A number of performance indicators are based on Self-Reported Measures, which are gleaned from either questionnaires, rating scales, interviews, focus-groups, or other methods requiring introspection on the part of the participant. These subjective measures are typically not logged continuously, but rather only once or a few times during the course of one study. The measures related to Self-Reported performance indicators could be the answers to each single question or the checks on the rating scales, while the sensors would be the questionnaires or rating scales themselves. It is more difficult to make a meaningful distinction between measure and sensor for semi- and unstructured interviews and especially for focus groups.

Subjective data, e.g. on acceptance and trust of a system or function, can provide valuable performance indicators, and in particular such data can be related to function usage in cases where this is within the control of the operator. Consideration should be given to tracking such acceptance and trust over time, as the levels may change with experience of the function.

In the matrix only a small number of Self-Reported Measures are included, which are those that are necessary for the computation of a performance indicator that is not solely based on self-reported measures, like for example “deviation from intended lane” or “rate of errors”.

5.3.5 Situational variables

Situational Variables can be logged like Direct Measures or computed like Derived Measures. They can also be self-reported and they can correspond to events. Their commonality is that they can be used as a differentiation basis for other performance indicators, in order to allow for a more detailed analysis. It might, for example, be of interest to compare certain performance indicators in different weather or lighting conditions, on different road types, or for different friction conditions. These Situational Variables are included in the performance indicator (PI) matrix in the measures table, but they are not linked to any specific performance indicator. In principle all kinds of measures can be used as Situational Variables, such as when analyses are performed for different speed intervals.

Data on Situational Variables are essential to collect, since it helps to establish important control factors that are needed when analysing the effects observed in the FOT. Ideally a lot of in-depth data is collected, such as:

- Video data (video processing takes a lot of time so this should be automated as much as possible)
- Questionnaires & travel diaries (at certain points during the test, too frequent interventions will disturb the naturalistic approach)
- Data on surroundings (such as surrounding vehicles – headways – and traffic state)
- Meta data; description of data and tests for evaluation, such as who drove the vehicle, which functions were studied in a particular test drive, circumstances when driving, other functions in the vehicle, date and time of the test, etc.
- Audio data; e.g. to give test drivers the options to tell what happened and what their experiences are, to make it as easy as possible for them – e.g. using a voice memo
- Logging of function states and messages (system state, e.g. on/off, what information is presented to the driver)

However, collecting all these data is costly and makes the analyses time-consuming. A possible solution is to collect in-depth data for part of the FOT (e.g. for part of the vehicles).

5.4 The PI-Measures-Sensors matrix

A matrix was developed that in one table contains performance indicators covering different aspects of research questions that might be addressed in an FOT (see the annex of FESTA Deliverable 2.1). These performance indicators are described with respect to different categories. For each performance indicator, the measures on which it is based are listed.

All these measures are then described in another table. Different categories are provided for description, where some are reserved for Direct Measures, others for Derived Measures and for events. Each Direct Measure points to a sensor from which the measure can be read. As mentioned above, for certain measures different sensors can be used. In this case, each of those is described as a separate measure.

A link is made between the PIs and the measures table by indicating for each performance indicator which measure is needed to compute it. In this way, when the hypotheses have been generated, it should be possible to pick the appropriate performance indicator and from there proceed via the pointers to the necessary measures and from there to the sensors. If several sensors can provide the same measures choices can be made due to budget limitations, sensor limitations or other restrictions.

Presently most measures for Self-Reported performance indicator are not included in the matrix. Instead, a direct reference is made to the appropriate questionnaire, rating scale or method needed to obtain this PI. For correct deployment of the recommended method, the user is directed to the instructions for this particular method.

Measures that describe driver characteristics are not included in the matrix itself, but in the annex to the matrix. In this annex, it is explained which instruments could be used to assess different aspects of driver characteristics (FESTA D2.1). The characteristics covered in this document are usually stable over a longer period of time.

This matrix is not meant to be exhaustive; it is only an aid for selecting performance indicators, measures and sensors. It should by no means be regarded as being limited to the performance indicators or measures entered now, and users are encouraged to expand the matrix during the course of their FOTs. Further instructions on how to work with the matrix are provided in FESTA Deliverable D2.1.

5.5 Performance Indicators per impact area

The performance indicators are split into different sub-groups, depending on which area of the traffic system they are concerned with.

5.5.1 Indicators of driving performance and safety

Driving performance is discussed and analysed in relation to traffic safety. Given that accidents are usually multicausal, the desired set of indicators should cover a number of factors. Otherwise any FOT is likely to miss essential information that is required to produce reliable and valid results.

Traffic safety is regarded as a multiplication of three orthogonal factors, namely exposure, accident risk and injury risk (Nilsson, 2004). The driver decision making and behaviour covers all these aspects. Typically, strategic decisions are highly relevant for exposure, tactical decisions for the risk of a collision, and operational decisions for the risk of injuries (Michon, 1985). Consequently, an FOT should cover all these aspects, because it is essential to cover the driver tasks and driver behaviour widely, and include decisions like whether to use the vehicle at all, route planning before the trip, timing of the trip, etc. However, as those decisions often lay outside what can be influenced with available countermeasures, the main focus is usually on driving performance while actually driving a vehicle.

The most common approach to traffic safety in NDS studies is to contrast driver behaviour: normal driving and the sequence of events leading up to conflict, situations (i.e. near-crashes or crashes).

In summary, an *indicator of driving performance* is a behavioural variable which *indicates* the quality of the behaviour in respect to road safety. The behaviour is measured directly from driver (e.g. frequency of glance to given object) or indirectly from the vehicle (e.g. speed).

5.5.2 Indicators of system performance and influence on driver behaviour

In this part indicators were developed that describe the actual performance of the system to be tested. These indicators are mostly related to both *safety* and *acceptability*. Here the focus is directed at the question whether the system actually functions the way it is meant to under realistic conditions. False alarms and misses are obvious indicators in this regard. Relations exist with indicators of acceptance and trust, which examine the subjective opinion of the participants on how the system worked.

Furthermore, indicators that describe the influence of the system on the driver and the interaction between system and driver are described. They will enable assessing the driver's willingness to use the system in various situational contexts. They will also contribute to the identification of potential misuses of the system leading to incidents or conflicts. In a longitudinal perspective, they will also contribute to an analysis of the learning and appropriation phases.

5.5.2.1 The intrinsic performance of the system

The first issue is the intrinsic performance of the system studied. It is related to the *precision and the reliability of the system*. Does the system perform as expected? In this case we need indicators signalling any deviations, such as false alarms and misses, but also indicators about the context in which these deviations occur. Ideally, the origin of the deviation should also be identified. The identification of false alarms or misses may be based on automated sensors or may require a video recording of the driving scene. For example, in the French LAVIA (ISA) project, loss of the recommended or target speed was automatically recorded while mismatching between the target speed and the posted speed limit was identified on the basis of a video recording of the driving scene.

The intrinsic performance of the system should be distinguished from the operational envelope of the system (i.e. the use cases[FW] for which the system was designed to work).

This is important when assessing the opinion on the performance of the system: when asking the driver to assess the system performance, the limits of the system operation should be differentiated from system deviations. Two main indicators related to the operational envelope are: 1) availability of the system over driving time (Percentage of the driving time the system is available, e.g. some system are only available above a certain speed, for special road characteristics, etc.); and 2) frequency of take-over requests (the system is active but not able to provide assistance due to system limits, e.g. for ACC maximum brake rate is limited).

Both intrinsic performance and competence envelope are assumed to play a role for the drivers' opinion on the system.

5.5.2.2 Modes of drivers' interaction with the system

The second issue is the driver's interaction with the system. This goes beyond the analysis of overall driving performance when using support systems: in fact 1) it is examined how drivers use and interact with the system; and 2) it is examined how this interaction may affect driving behaviour and performance.

5.5.2.3 How drivers use and interact with the system

Some support systems require/enable the driver to activate/deactivate the system, to override the system, to select one system among other systems available, to select or to register some vehicle-following or speed thresholds, and so on. In other words, using a system implies the application of a number of procedures, and these procedures should be registered and analysed. This is the case for systems such as speed limiters, cruise control, adaptive cruise control or navigation systems for example. These procedures may be classified as the driver's direct or indirect interventions, depending on whether they are applied through vehicle controls (brake or accelerator) or through system controls. As for the indicators of system performance the situational context should be taken into account. This is important for identifying potential misuses of the system leading to incidents or conflicts as described above. In a longitudinal perspective, these indicators will also contribute to an analysis of the evolution of system usage from the learning and appropriation phases to the integration phase. Furthermore, the frequency with which the system "interferes" with the driver's activity has to be assessed. For example, when driving with a speed limiter, how often is the system "active", that is, effectively limiting the vehicle speed.

5.5.2.4 How this interaction may affect driving behaviour and performance

For analysing the effect of the driver's interaction with the system on driving behaviour and performance various levels of analysis could be employed, depending on the desired level of granularity of analysis. Obviously, this granularity depends on the recording means available as well as on the time required for performing such analyses. For example, studying changes in glance behaviour requires video recordings and is time consuming.

For an analysis of behavioural changes at a more global level, synthetic indicators should be conceived. These indicators are assumed to reflect changes at the tactical or strategic level of the driving task. Indicators such as "lane occupancy" and "frequency of lane change" are

often used to assess changes at the tactical level. Changes at the strategic level could be reflected by changes in the itinerary chosen or changes in driving time.

Recommendations:

- Classify the support systems by type and level of interaction implied by their use;
- Classify the performance indicators according to the level of granularity of analysis that they permit;
- Classify the performance indicators according to the means and time required for collecting and analysing them.

5.5.3 Performance indicators of environmental aspects

Exhaust emissions include many different substances like: HC, CO, NO_x, PM, CO_2, CH_4, NMHC, Pb, SO_2, N_2O and NH_3. Greenhouse gases – CO_2, CH_4 and N_2O – represent the same society cost anywhere, while costs for other substances depend on the geographical position.

There are two alternatives for quantifying exhaust emissions: measured exhaust emissions or calculated. For measurements there are still two alternatives: on board or in the laboratory. The laboratory alternative demands use of logged driving patterns. Because of the high complexity and costs for measurements of exhaust emissions, in practice, calculated emissions is in most cases the only reasonable alternative.

Models for exhaust emissions in general include three parts: cold start emissions; hot engine emissions and evaporative emissions. The following formula is a rough description of an exhaust emission model:

Σ(Traffic activity) x (Emission factor)=Total emissions

Traffic activity data include at least: *mileage* and *engine starts*. Hot emission factors for one vehicle are functions of the driving pattern and vehicle parameters. Cold start emission factors are functions of the engine start temperature, trip length and average speed. Evaporative emissions are to a large extent a function of fuel quality and fuel tank temperature variations.

Models on a micro level, including engine simulation, should in principle be able to describe most ICT functions. This is not the case for models on a macro level in general. Micro models are often used for emission factor estimation and macro models for total emission estimations.

The conclusion about what to include as performance indicators would then be: exhaust emissions or measures with high correlation to exhaust emissions.

5.5.4 Indicators of traffic efficiency

The efficiency of a traffic system can be measured as, for example, traffic flow, speed and density in relation to the optimum levels of these properties given the traffic demand and the physical properties of the road network.

A combination of FOTs and traffic modelling is required to allow estimation of traffic efficiency impacts of the tested technologies. A schematic picture of the proposed methodology is shown in Figure 5.1.

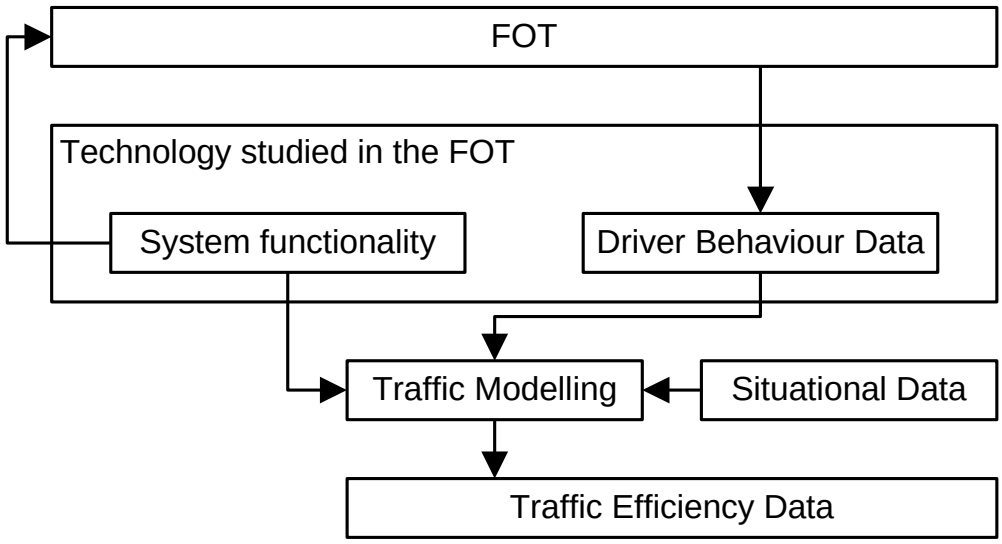


Figure 5.1: FESTA Traffic efficiency estimation based on FOT results

Driver behaviour data are based on the data collected in the FOT . These driver behaviour data will, together with the system functionality[11] of the tested technology, be used as input to traffic modelling in order to aggregate the individual driver/vehicle impact on traffic efficiency effects. This requires that both driver/vehicle data of equipped vehicles and properties of the traffic system that the vehicles have driven in (henceforth referred to as Situational Variables[12]) are collected in the FOT.

The driver behaviour data required in order to estimate traffic efficiency for any type of FOT system are specified in terms of performance indicators and Measures and included in the attached matrix. These data (along with the Situational Variables, which can be found in the Measures Table in the Annex of Deliverable D2.1 should be ascertained for the baseline case (non-equipped vehicle) and for equipped vehicles, so that comparisons can be made between the two).

The appropriate traffic modelling approach will differ depending on which type of driving tasks supported by the considered technology. Michon's (1985) hierarchical driving model

[11] System functionality refers to the way in which the tested FOT system works. Information on when and how the system operates can be used to create parameters for the models developed.

[12] Situational Variables are not necessarily directly relevant for Performance Indicators or Derived Measures, but must also be measured or recorded as they provide key background information that complements the driver behaviour data and is sometimes needed to derive the driver behaviour data. Examples include light conditions and road type.

can be applied to select a traffic modelling approach. To model systems that support tactical or operational driving tasks, it is appropriate to apply a traffic microsimulation model. A microsimulation model considers individual vehicles in the traffic stream and models vehicle-vehicle interactions and vehicle-infrastructure interactions. To model systems that support strategic and some types of tactical driving tasks it is appropriate to apply a traffic simulation model. A mesoscopic model considers individual vehicles but model their movements and interactions with a lower level of detail than microscopic models.

It is advisable to study traffic efficiency for a series of scenarios with varying levels of traffic penetration of the tested systems. The systems should also be studied in representative traffic volumes. This is achieved straightforwardly by running the traffic simulation model with different inputs. The situational data will also contribute to the differences between the scenarios (both measured and modelled).

Outputs from the traffic models will be used to make comparisons of traffic efficiency for the studied scenarios. Example outputs of interest are traditional quality of service and traffic efficiency indicators such as speed, travel time, and queue length.

5.5.5 Acceptance and trust

Acceptability indicates the degree of approval of a technology by the users. It depends on whether the technology can satisfy the needs and expectations of its users and potential stakeholders. Within the framework of introducing new technologies, acceptability relates to social and individual aspects as well.

Regarding the dimension of “Acceptance and Trust”, the following subjective PIs should be focused on during FOTs:

Ex-ante usefulness (level of usefulness perceived by the user prior to usage): before using a system, what are the dimensions of usefulness that occur to the future user immediately? What are the benefits he expects from using the system? **Ex-post usefulness** (level of usefulness perceived by the user after practice with the system): after a first use of a system, what are the user’s impressions regarding the system’s benefits. Ex-post usefulness is to be analysed in relation to the statements of the indicator on “ex-ante usefulness”. The reactions to both indicators will give useful information for system acceptance. The measurement of these two indicators can be operationalised via self-designed questionnaires, based on established methodological approaches (see Nielsen, 1993; Grudin, 1992). A qualitative approach like a Focus Group with a formalized protocol and individual in-depth interviews is also appropriate. The **Observed rate of use** of the system or of specific system parts represents an additional indicator for system acceptance and perceived usefulness. **Perceived system consequences** (perception of positive or negative consequences of system's use) is another key indicator for system performance: the user expresses his/her impressions and attitudes regarding the potential consequences when using the system, which can be positive as well as negative. These impressions can best be collected via an interview and be exploited in Focus Groups, which have the advantage of group dynamics that can provide additional information on the subjective norm. Construction of standardised questionnaires is possible as well (for methodological

background on this indicator, see Featherman and Pavlou, 2003). **Motivation** (level of motivation/impetus to use system) should be connected with the indicator **Behavioural intention** (level of intention to use system). Both indicators can be investigated best via self-designed questionnaires based on established methodological findings (see Armstrong, 1999; Ajzen and Fishbein, 1980). The **Response to perceived social control/response to perceived societal expectations** indicates the impact of perceived social control of the user's behaviour. This indicator is a more sociological one, which should give an indication whether the user feels a social benefit (for example, social recognition) when using the system, or on the contrary, that he hesitates to use the system because he fears social disapproval when using the system (see Castells, 2002). **Usability/level of perceived usability** concerns the aspects of the user's general capacity to interact with the system (including installation and maintenance issues, see Grudin, 1992; Shakel & Richardson, 1991). For these indicators, the combination of in-depth interviews, Focus Groups and self-designed questionnaires based on established methodology is recommended.

5.5.6 Driver characteristics

Even though driver characteristics are not performance indicators in themselves, they are important as Situational Variables, which is why they are included in this section. The focus here is on describing the drivers that participate in the study, as compared to selecting drivers based on certain characteristics, which is treated in Chapter 6. Drivers differ on a large variety of characteristics, which may all have an influence on how they drive and use different systems and services. These differences may be important to take into account when planning an FOT. Four categories of driver characteristics may be distinguished:

- Demographic characteristics: gender, age, country, educational level, income, socio-cultural background, life and living situation, etc.
- Driving experience, and driving situation and motivation: experience in years and in mileage, professional, tourist, with or without passengers and children etc.
- Personality traits and physical characteristics: sensation seeking, locus of control, cognitive skills, physical impairments or weaknesses, etc.
- Attitudes and intentions: attitudes towards safety, environment, technology etc.

Studies often focus on characteristics of individual drivers. However, drivers are not alone on the road. There are other road users and there may be passengers in the vehicle, which may influence the driver's behaviour.

There are several different reasons for considering driver characteristics:

- To make sure that the sample of drivers is representative of the target population.
- To explain the outcomes of the FOT.
- To improve systems and services, taking into account differences between drivers.

Driver characteristics may play different roles in FOTs:

- Characteristics of drivers possessed before the FOT may play a role in how they behave in traffic during the FOT].
- Although some characteristics are stable, other ones may change when using a system or service in the FOT. Attitudes may change radically before and after using a system for a longer period of time.

In general it is useful in an FOT to gather as many characteristics of drivers as practically possible. Even if no specific impacts are expected of certain characteristics, some outcomes may be explained better with more knowledge about the participants. A minimum set of data such as age, gender, income group and educational level is easy to gather from participants.

Next, information is needed about driving experience. Usually this is measured by means of self-reports. The amount of practice, i.e. the mileage of an individual driver can be collected by asking the subject for an estimation of his/her overall mileage since licensing or the current mileage per year. However, beware that these self-reports are not very reliable.

For further understanding of driver behaviour one may consider to use questionnaires on attitudes, driving behaviour and personality traits. A well-known questionnaire about (self-reported) driving behaviour is the Driver Behaviour Questionnaire. Some widely used personality tests are the Five Factor Model (FFM) test and the Traffic Locus of Control test(T-LOC). Special attention may be given to the personality trait of sensation seeking, which is correlated with risky driving. The Sensation Seeking Scale (SSS) measures this trait. These questionnaires are available in many different languages, but they are not always standardised, and cultural differences may play a role. Personality traits are very easy to measure, just by administering a short questionnaire. However, the concepts and interrelations of factors are very complex, and results should be treated with caution.

When evaluating the acceptance and use of new systems in the vehicle, drivers' acceptability of technology is important. Both social and practical aspects play a role. Technology acceptance has different dimensions, such as diffusion of technology in the drivers' reference group, the intention of using the technology, and the context of use (both personal and interpersonal). Measuring acceptability can be realized via (existing) standardised questionnaires, in-depth interviews before and after "use" (driving), and focus groups.

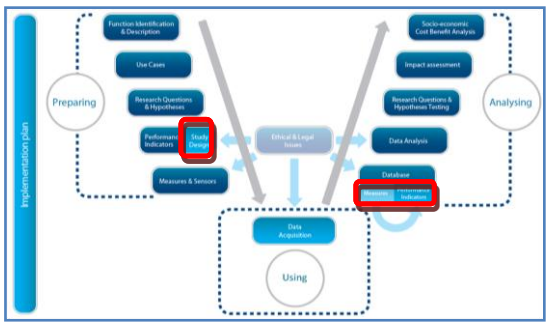
5.6 Iteration

When the performance indicators have been defined, it is recommended to re-check whether these indicators are indeed capable to test the hypotheses defined earlier, and if necessary to adjust the hypotheses or the indicators. Available resources will play a major role in determining which performance indicators to use. It is also necessary to look forward in the FESTA chain, and to consider data storage and analysis. If a large number of performance indicators has been selected, or if the performance indicators require a huge amount of data to be collected, considerations about data collection and storage capacity

come into play as well as the question how to analyse those data. For example video data require a large capacity and ample resources to analyse them.

If there are foreseeable problems with this it may be necessary to limit the amount of performance indicators.

6. Experimental procedures



This section of the handbook provides guidance on the overall experimental design of FOTs[FW] in order to ensure experimental rigour and scientific quality. The first section — Participants — provides advice on participant selection, including demographics, driving experience, personality and attitudes, along with consideration of sample size. The second section – Study design – provides guidance of the formulation of hypothesis[FW], experimental design and possible confounds. The third section – Experimental environment– suggests how the road environment (road type, weather conditions etc.) plays a part in the design of an FOT and the subsequent data analysis. In the fourth section piloting is explained and in the last section the methods of controlled and semi-controlled testing is explained.

6.1 Participants

6.1.1 Characteristics

Depending upon the research questions[FW], there is often a need to select a particular group of participants for inclusion in the FOT and ensure that this group is in some way representative of those drivers who will ultimately interact with the system.

The types of variables that should be taken into account include:

- Demographics variables, such as age, gender, social economic variables, and permanent or temporary driver impairments
- Driving experience, in general but also experience with various systems[FW], accident history and the usual time of driving and roads used
- Personality and attitudes.

The first of these two variables are relatively easy to measure, using questionnaires. The data are objective and can be verified by the experimenter. Personality and attitudes, however, deserve more attention as there are a number of different ways in which one can evaluate these (see 5.5.6). FOTs may incorporate a battery of psychometric measures. Such measures are generally included in order to relate psychological factors to driving behaviour. Since drivers exhibiting certain traits or attitudes are known to engage in riskier driving behaviours, it would seem important that systems under investigation in FOTs are trialled amongst a range of drivers to ensure that the systems work for those who need it most.

Personality aspects that may be taken into account are:

- Sensation seekers, who tend to drive more recklessly
- Locus of control: drivers with an internal locus of control will continue to maintain direct involvement with the driving task choosing to rely on their own skills, whilst

those with an external locus of control may be more likely to rely on the system and surrender involvement in the driving task

- Drivers' attitudes towards road safety issues.

Personality and attitudes are known to affect the ways in which drivers interact with systems, and it may therefore be of interest to preselect certain personality types in much the same was as one would sample e.g. young males, or elderly drivers to a particular trial.

Recruiting on a personality/attitude base will ensure that a system is tested on a broad range of drivers who may interact with the system very differently. Recruiting on a personality/attitude base may be appropriate since these are likely to influence behaviour directly. Variations in beliefs are likely to explain differences in driver behaviour and system use. Before beginning recruitment for any FOT. researchers must consider the relationship between individual differences and the behaviour which the system is seeking to influence.

In addition to selecting drivers, personality and attitudes can also be used as covariates in analysis in order to identify several differences in driver behaviour and system use between groups. It is not imperative that FOTs base their recruitment on such measures. However, their inclusion within the experimental design provides useful insight into the manner in which individual characteristics influence behavioural adaptation to new systems.

Before deciding to recruit on a personality/attitudinal base, researchers should consider that, when tiding the inclusion criteria for any study, it is inevitable that there will be a progressive shrinking of the research participant *population*. It may therefore be necessary to screen a large number of drivers in order to recruit a relatively small number of participants with the appropriate characteristics, particularly since certain individuals will be less inclined to volunteer to trial certain systems. For example, since speeding represents a thrill seeking behaviour, high sensation seekers may be less likely to volunteer to participate in an ISA trial. Inevitably selecting participants on additional measures such as these will increase the burden associated with the recruitment phase of any FOT.

6.1.2 Sample size and power analysis

FOT studies should be able to assess the functionality of the ICT systems and their impact on the driver behaviour, traffic safety, environment, etc. When the chosen sample size is too small, it is difficult to statistically prove effects of the system that are actually there. With very large sample sizes the chance of finding an effect increases. However, there are two major drawbacks on using very large sample sizes:

- Every driver/participant needs a car equipped with the system and with a data logging system, which is expensive.
- Small effects which are statistically significant might be found, but they might not be relevant when looking at power effect.

The appropriate sample size for an FOT depends on a number of choices that have to be made in the final setup[13]. These are, for instance, the number of ICT systems that are going to be tested and the choice of a between-subjects (two separate groups of drivers with and without an ICT system, but always with data logger) or a within-subjects design (each participant drives a certain amount of time with and without the ICT system).

In order to ensure that the chosen sample size is representative for the behaviour of a group of drivers and that it is possible to statistically prove effects that are there, power analysis is needed to calculate the desirable sample size. This power analysis is based on a number of assumptions:

- Suppose an FOT is based on a between subjects design, such that different groups of drivers each drive with a different system. Or at least one group with an ICT system and one group without and ICT system
- The power is 80 %, indicating the chance of statistically proving a difference between the groups when it is there (i.e. a chance of 20 % of failing to prove it)
- The alpha level is 5 % (i.e. the chance of falsely finding a significant effect)
- Two-tailed testing, because we have no reason to assume that either one of the groups performs better/worse than the other

The effect size is 0.2, which is typical for a small effect that can be expected in an FOT with a lot of disturbing factors compared to more experimental test set-ups. An effect size of 0.5 is typical for a medium size effect. EuroFOTanalysis has indicated that it is more effective to increase the number of drivers than to extend the time period of data collection (Jamson et al., 2009).

[13] For further information on how to choose the sample size, the reader should refer to FESTA Deliverable D.2.4.

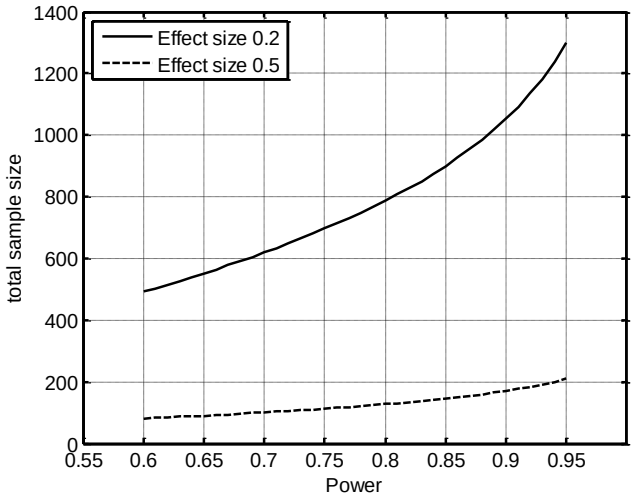


Figure 6.1:Total sample size as a function of the statistical power and the effect size (2-sided test, alpha = 0.05, independent variables).

Figure 6.1: shows that a total sample size of 800 (i.e. two groups of 400) drivers would be needed to be able to statistically prove small size effects between the two groups. The groups are relatively large to compensate for the relatively high number of disturbing factors when trying to find effects in real traffic. If we expect medium size effects groups of only 75 drivers would be sufficient. If a within subjects design is chosen, one group of 400 drivers would be sufficient to test both the without and with system conditions.

In practice, recruiting the specified target sample may turn out very difficult. In order to compare results between countries, it would be ideal to have the same equipment in all countries as well as the same group of drivers. This can however prove to be difficult because of the penetration rate of different makes of car in the vehicle fleet of different countries and drivers and their 'national driving styles' being different.

6.2 Study design

6.2.1 Hypothesis formulation

Hypothesis formulation is described in section 4.2.4 and subsections.

As a general rule, research practice proceeds in the following way:

1. Formulation of the hypothesis
2. Testing the hypothesis
3. Acceptance or rejection of the hypothesis
4. Replication of the results or (in the case of rejection) refinement of the hypothesis

A hypothesis is a specific statement which can be tested with statistical means by analysing measures and performance indicators. It is a tentative explanation for certain behaviours,

phenomena, or events that will occur. It is essential for an FOT to be designed with clear hypotheses in mind in order to aid the interpretation of the results.

In formulating a hypothesis, consideration should be given to the variables under scrutiny. It is vital that the variables collected in an FOT allow the researcher to accept or reject their hypotheses. To do this, both the independent and dependent variables should be well defined at the start of the FOT. The independent variable is one which can be manipulated by the researcher. As the researcher changes the independent variable, he or she records what happens using dependent variable(s). The resulting value of the dependent variable is caused by and depends on the value of the independent variable. Other variables, known as controlled or constant variables are those which a researcher wants to remain constant and thus should observe them as carefully as the dependent variables. Most studies have more than one controlled variable.

6.2.2 Experimental design

The two basic types of experimental designs are *within subject design* (this is sometimes also referred to as crossed design) and *between subject design* (this is sometimes also referred to as nested design). FOTs also need to contain a control condition, in which subjects do not get any treatment. This condition is meant to serve as the baseline: This is how drivers behave in case there is no treatment or no experimental manipulation at all.

6.2.2.1 Within subject design

In a within subject design, each subject encounters every level of treatment or experiences all experimental manipulations. For example in an FOT evaluating navigation systems, every subject drives for some time with (experimental condition) and for some time without (control condition) the system. In this specific case, one half of the subjects would start with the control condition and then switch to the navigation (experimental) condition and half of the subjects would do this vice versa.

This type of design has two advantages: (1) fewer subjects are needed compared to a between subject design, and (2) is more likely to find a significant effect, given the effects are real. The power of a within subject design is higher than in a between subject design. This is related to the reduction in error variance, since there are no individual differences connected to differences in treatment measures. A disadvantage is the risk for carry-over effects, which means that if a subject experiences one condition, this may affect driving in the other condition.

6.2.2.2 Between subject design

In a between subjects design, each subject participates in one experimental (or control) condition. The major distinguishing feature is that each subject has a single score (with or without the system). Note that the single score can still consist of driving on various types of roads, during long periods of time or different types of driving behaviour, workload and comfort.

The advantage here is that carry-over effects are not a problem, as individuals are measured only once in every condition. The total number of subjects needed to discover effects is

greater than with within subject designs. The more treatments in a between subject design, the more subjects are needed altogether. In order to limit the confounding effects due to individual differences in a between subject design, one should either use random assignment, in which the assignment of what subject is exposed to what treatment is done randomly or use matching groups (also called matched pairs), in which one also has to make sure that different groups are comparable with respect to pre-selected characteristics, such as gender and age. In order to do this, one needs to identify the variables that one wants to match across the groups, and measure the matching variable for each participant and one needs to assign the participants to groups by means of a restricted random assignment to ensure a balance between groups. Also, one needs to keep the variable constant or restrict its range. This will reduce differences within each group and therefore reduce within treatment variability.

The main drawback with the matched pairs design is in the sampling process. As the number of characteristics that require matching increases, so a correspondingly large sample pool will be required to allow adequate matching to be possible. A further problem is that this design assumes that the researcher actually knows what extraneous factors need to be controlled for, i.e. matched — and in some circumstances this may not always be the case.

6.2.2.3 Longitudinal and Cross-Sectional Designs

One question an FOT may have to answer is whether an effect of a treatment (e.g. driving with a system) changes over time? To investigate this, longitudinal or cross-sectional designs can be employed. While longitudinal surveys of this type can be very useful they do not provide an answer to the questions concerning why the changes may or may not have occurred. If things like that are measured in FOTs , one should already have a clear idea why a positive effect may disappear after a while. This could for instance be such factors as risk compensation (because the systems warn you, you can drive until you are warned).

One of the difficulties with longitudinal studies is that it is hard to keep subjects motivated during the entire study period, or people may move, or become ill. Because of these difficulties other methods for investigating changes over time have been developed and the cross-sectional design offers an alternative.

The cross-sectional design looks at changes over time by taking a number of cross-sections of the population at the same instant in time. This is obviously quicker and less costly than a longitudinal study, and there is a lower chance of actually ‘losing’ participants during the run of the experiment. On the other hand, a main drawback with the cross-sectional study is related to the previous experiences of the participants and how this might have an impact on the findings.

6.2.2.4 Baseline and treatment period

The baseline period is often squeezed in the project and is quite short, especially in relation to the treatment period. Ideally, the two would be equal lengths so that there is the opportunity in the baseline period for the same variations to occur that may occur in the

treatment phase (such as seasonal effects, see 6.3.6). The more data available the more robust the results are.

6.2.3 Threats to validity: confounds and other interfering effects

As a general rule, the results of an empirical study should allow a clear decision if the hypothesized relationships between variables exist or not, i.e. if the hypotheses can be accepted or has to be rejected. In the best case, the researcher is able to attribute the changes he/she observed at the dependent variable without any doubts to the manipulation of the independent variable. The internal validity of an experimental or quasi-experimental study describes the extent to which this inference is unequivocally possible because the study has been designed in a way that alternative explanations for the effects are implausible or can be excluded. The internal validity of a study increases to the extent to which such alternative explanations can be ruled out. In the literature these factors are also described as confounded variables which need to be controlled by appropriate measures right from the beginning of a study.

In the literature several interfering effects have been described which interfere with the effect of an independent variable on a dependent variable and contribute to a decrease of internal validity if they are not controlled by measures implemented in the experimental design. The following effects constitute threats for internal validity of FOTs:

- **History:** Unplanned events unrelated to the study might have an effect on the correlation between independent and dependent variables. For example, during the performance of an FOT an important paragraph of the road code might be changed (e.g. new speed limits for certain road categories) which is accompanied by increased police surveillance activities.

- **Maturation:** Mainly effects due to experience and learning which affect the dependent variable and are (in long-term studies) erroneously attributed to the independent variable.

- **Testing:** If the behaviour of interest is sampled at different times there might be a biasing effect from the number of times, e.g. by becoming more familiar with the test situation[FW]. For FOTs this might become relevant if subjects are tested at different times over the course of the study but not if their behaviour is sampled continuously and more or less unobtrusively.

- **Selection:** In general, the participation in an FOT is voluntary which means that the strategy of recruiting subjects can have a biasing effect. For example, to offer a certain amount of money (e.g. 500 Euro) as compensation for the effort caused by completely finalising the study might be an incentive for participants with a low income whereas it might insult people with a very high income.

- **Drop-out:** During the run of an FOT one has to take into account that not all subjects will finalise their participation as planned. However, this drop-out can have a biasing effect on the results of an FOT if the subjects who quit early differ systematically

from those who finalise as planned with regard to relevant characteristics (e. g. socio-economic status, age, gender etc.).

- **Experimenter-bias:** Effects on the dependent variable which result from the social interaction between the experimenter and the subjects which might occur, for example, if at the beginning of an FOT the experimenter explains the system functions very carefully to some subjects due to sympathy whereas he is careless with this at some others.

6.3 Experimental environment

The experimental environment is a critical element within an FOT, since it will determine the data that is collected and the ability to fulfil the objectives of the FOT. In general, environmental factors can be treated in several different ways, including

- Explicitly included in an FOT because there is a particular interest in data connected to that environmental factor (e.g. motorway routes for lane departure warnings)
- Explicitly included in an FOT because these environmental factors are part of the range occurring within a normal driving scenario (e.g. night time driving)
- Measured scientifically so that the data relating to that environmental factor can be included within post trial data analysis (e.g. vehicle headways)
- Recorded (in varying levels of detail), so that portions of data can be excluded from analysis (e.g. heavy rain, where all or some of the data from a particular day may be discarded; or overtaking manoeuvres where short periods of data within a larger set are discounted during a study of steady following behaviour).

6.3.1 Geographical location

In line with above, the geographic location can be chosen because it is representative of the intended area of use for a vehicle/system (e.g. predominantly motorway environments). Alternatively, the geographic area can be chosen because it displays the characteristics needed to collect the specific data the srudy is interested in during the FOT (e.g. the choice of mountainous and/or northern European environments in order to collect data on the use of systems in cold environments).

The population within a particular geographical location may affect the running of the FOT . For example, certain cultural issues, population characteristics, car ownership, use of new technologies, and language issues may be apparent. In addition the characteristics pertaining to the road and prevailing traffic may be of importance, including:

- Road type and localities present
- Traffic patterns, such as types of journeys (e.g. commuter or tourist travel), traffic flow, traffic density, vehicle types, and frequency and sophistication of journeys

- Other transport options, the availability and costs and the inducement or penalties to encourage particular transport mode choices
- Legal regulatory and enforcement environment, such as speed limits, levels of enforcement of traffic regulations (e.g. speed cameras), penalties for traffic or other violations, standardisation (e.g. compliance of road signs with international standards).

The geographical location may also have implications with regards to technical and other study issues, including infrastructure and data communication issues such as:

- Network/beacon infrastructure for vehicle-infrastructure communication
- Network coverage/reliability for telecommunications, especially if automatic over-the-air data transmission is used instead of manual data download
- Localised GPS coverage issues (e.g. urban canyons, foliage cover)
- Logistical issues, both in the validation and the experimentation phase safe and secure access to infrastructure equipment should be ensured for validation of the functions(especially in case of cooperative systems), for data download (if remote access is not available) and maintenance. As well target vehicles should be accessed for data download (if data is not being transmitted over the air) and for maintenance.
- The availability and quality (resolution, scope and depth of content) of electronic maps that can integrate vehicle location for situation evaluation. Moreover, in case of complex functions and especially for cooperative systems, high accuracy maps may be required in order to implement these functions.
- Availability of other data, e.g. from the police, highway authorities, fleet operators, maintenance personnel.

The most important point in relation to the geographical area is that it must be chosen based specifically on the *objectives* of the particular FOT, and in particular, in relation to the validity of the data that is being collected. There are two overall considerations:

- Is it needed to consider a particular geographical aspect because it is relevant to the types of vehicles and or systems being studied?
- Does a geographical aspect need to be considered to ensure that the results obtained can be generalised to the wider 'population' of interest (i.e. external validity)?

The starting point is to consider the overall objectives of the FOT , including the types of cars and systems that will be incorporated into the trial. The second major consideration is that of generalisation of the results. In particular it is necessary to ensure that geographical aspects are included to ensure that the data collected during a specific FOT can be generalised to the wider population of interest. The third factor to consider is whether the

geographical factor is of particular interest in terms of data analysis. If it is desirable to analyse results according the presence or absence of a particular factor, then the geographical environment(s) must include that factor (and possibly variation thereof).

Finally, it is important to note that the decision to collect data in a specific country might, due to legal requirements in the country, have an impact on how especially personal data later could be collected, handled and shared.

6.3.2 Road type

The road type is the environmental factor that perhaps has greatest dynamic influence on individual and collective driver behaviour, and hence impact on safety, mobility, traffic efficiency and the environment within an FOT. It is highly dependent on the geographic area, as discussed above.

The road type will encompass a number of variables which will influence driver use of systems, driver attitudes, driver behaviour, and driver outcomes. The FOT may want to include roads with specific characteristics, including:

- Surfaced or unsurfaced roads
- Minimum, average and maximum speeds of traffic
- Number of lanes, and presence of lane marking
- Visibility (of the environment and other traffic)
- The types of manoeuvres that a driver will need to undertake (e.g. stopping at traffic lights, or overtaking manoeuvres)
- Typical vehicular headways
- Presence of safety features such as rumble strips or speed cameras.

Three main categories of road should be differentiated:

- Urban
- Rural
- Motorway

Note that road classifications differ in different countries and there is no standard European classification. Ideally a map and a database of the region of deployment of the FOT should be established in order to reduce the time needed afterwards for collecting this type of data (on the basis of the video recording of the road scene). An electronic map containing at least the type of roads and the speed limits in force (and location of speed cameras) would greatly facilitate the task.

6.3.3 Traffic conditions and interactions with other road users

Traffic conditions and interactions with other road users are important considerations. A distinction needs to be made between:

1. Traffic conditions in a general sense, which characterize a general level of constraints and which, in the same manner as the infrastructure zones, define the driving environment
2. Other road users and their behaviour, which characterize an individual level of interaction between the driver and one or more other road users in the driver's immediate proximity.

The traffic, as a general and contextual entity, can be characterized using several dimensions, for example:

- Density: expressed in terms of the number of vehicles travelling in a given space;
- Stability: this can be within a traffic stream (in which case it is expressed in terms of the frequency of speed variations on a traffic lane in a given unit of time) or between different traffic streams (in which case it is expressed in terms of the frequency of lane changes in a given unit of time)
- Speed: the average speed of traffic
- Composition: types of vehicle (light vehicle, heavy vehicle, van, motorcycle) and their relative proportions in a given traffic stream.

The interactions at individual level between the driver and one or more other road users in the immediate vicinity can also be characterized using several dimensions:

- The category to which they belong (light vehicle, heavy vehicle, van, motorcycle, pedestrians)
- Their speed and acceleration (direction and rate)
- Their manoeuvres and behaviour (merging into the subject's lane or pulling out into a lane, merging from an entry slip road, braking, etc.).

Other characteristics to be taken into account are:

- Route choice
- Temporary road/traffic variables
- The traffic encountered
- Impact of road measures on driver behaviour
- Static and dynamic variables associated with the road

6.3.4 Roads to include

When setting up and running an FOT, it is necessary to consider the extent to which specific road types need to be incorporated into the trial and hence which participants need to be selected. The basic questions to consider are:

- Are specific road types needed to answer the research questions for that sample?
- Would any system of interest be used on a range of different road types?
- Is it expected driver behaviour (in terms of safety, mobility, traffic efficiency and environmental impact) to differ according to the road type they are travelling along?
- Is it needed to be able to compare results according to different road types?
- Is it needed to include specific road types in order to generalise the results to a wider population?
- Are interactions with other road users to be included in the analysis? If so video equipment needs to be installed.

By considering the above questions, one can determine whether a range of different road types are needed, or whether the FOT can concentrate on collecting data based on specific road types. In an FOT, the objective is usually to study the normal driver behaviour. This means that drivers should not be encouraged to change their normal routes.

6.3.5 Weather conditions

Weather conditions are hard to predict, control for, or measure accurately in an FOT. However, weather conditions and associated factors such as ambient lighting are relevant aspects for all FOTs, irrespective of the overall purpose of the study. A well designed FOT must consider a range of weather-related issues, with a view to including, targeting or excluding particular weather conditions. In order to include weather as an experimental variable within analysis, or to specifically include or exclude data for analysis, it is necessary to use a consistent taxonomy and definition of weather conditions.

Related to how weather factor are measured, is the level of accuracy that is employed in the measurement of weather factors, including location and time attributes. A further complication with weather factors is that it is often combinations of weather and other dynamic and static factors that have a practical impact on an individual driver or general traffic conditions within an FOT. Extreme weather conditions present a risk to FOTs because they often can't be predicted, and can make journeys impossible, prevent access to vehicles, or in the worst case can destroy equipment.

Data may be confounded due to abnormal weather, for example snowfall increasing driver headways and reducing traffic speed or bright sunshine causing glare on screens in vehicles, or momentary distraction to drivers.

There are several ways of potentially measuring weather conditions:

- In real-time using direct measurement of the factor, e.g. vehicle sensor to measure ambient temperature (which could then be used to link the use of features to outside temperature).
- Indirect real-time measurement using a surrogate sensor, e.g. recording the use of the windscreen wipers to indicate when it is raining.
- Subjective rating scales (completed by the driver or other) , e.g. a driver assessment of the degree of rainfall.
- Post-hoc data mapping – the use of weather records to estimate the weather conditions.
- Post-hoc analysis of video data by a trained data coder.

At a general level, there are four main considerations with regard to weather:

- Which weather conditions are relevant?
- Should they be 'designed in' or 'designed out' of the study?
- Do weather conditions of interest have a macro (e.g. a rainy day) or micro (e.g. reflected glare) level impact?
- What level of data is needed, and how is this obtained?

6.3.6 Time of day and seasonal effects

Temporal factors such as time of day, and seasonal effects have a considerable impact on the planning of FOTs, and the analysis of data. They can really cause problems for explaining the effects that are found (e.g. whether they are caused by the system under test or by seasonal circumstances). In contrast to the weather effects outlined above, the temporal factors can usually be predicted, and so it is usually easier to deal with the issues successfully. The main issues that have to do with the time of day, week, and seasonal variations are:

- Influence on driver state (e.g. sleepiness)
- Disruption caused by external events, for example school opening times
- Influence on traffic levels
- Other temporal influences on traffic
- Impact on vehicle occupants
- Glare
- Ambient light levels
- Seasonal confounding of data collection

- Influence on route choice
- Pragmatics to do with drivers work and life schedules
- Using time of day as a surrogate, for example, time of day can be used to specify or control for traffic levels or ambient light levels.

Time of day and seasonal effects are different to weather issues in several ways, including:

- Time of day and seasonal effects are much more predictable than weather conditions
- They are often proxies – i.e. not important in themselves, but important because they result in variation of a factor of interest (e.g. traffic levels, or level of the sun above the horizon)

These two factors mean that a greater emphasis should be placed on planning around relatively predictable time of day and seasonal effects, and considering their impact on the FOT. There are different ways to (partly) deal with seasonal effects: have a control group, or adjust the length of the test. The latter means that either there is a short time period for the FOT, so that baseline and treatment phase take place in the same season, or that the FOT is very long (more than a year), so that baseline and treatment phase include the same seasons.

6.4 Conducting a pilot study to test the evaluation process

A pilot study can be defined as a "small scale version, or trial run, done in preparation for the major study" (Polit et al., 2001); it goes before large-scale quantitative research and is very useful to test the research instruments, identify any performance problems and ensure a reasonable durability of the technology instruments adopted. Conducting a pilot study is a fundamental phase to get warning in advance about practical problems or difficulties that may affect the study and it is also necessary to prepare the deployment of the FOT and to support the design of the relevant tools for the evaluation process (Saad, 1997; Saad and Dionisio, 2007). This task should be performed early in the evaluation process. It represents an important step for the mobilisation and the dialogue between the various teams involved in the FOT and for promoting a common framework and consensus for the evaluation process.

The relevance of conducting a pilot study and the time required are often under-estimated. To better understand the importance of this step a list of general reasons for conducting a pilot study (for a wider overview, see Polit et al., 2001) is shown below:

- Developing and testing adequacy of research instruments;
- Assessing the feasibility of the full-scale study;
- Testing the research protocol;
- Testing whether the sampling frame and technique are effective;

- Verifying the likely success of proposed recruitment approach;
- Identifying logistical problems which might occur using proposed methods;
- Testing variability in outcome to help determining sample size;
- Collecting preliminary data;
- Verifying what resources (finance, staff) are needed for a planned study;
- Verifying the proposed data analysis techniques to uncover potential problems;
- Testing the research questions and research plan;
- Training the researchers , both in data analysis and in personal integrity issues.

Going more in detail, in FOTs these preliminary field tests have to deal with three main levels of analysis with specific objectives.

1. Obviously, the first preliminary field tests have to check *the technical functioning of the data collection* systems *in real driving* [situations](FW). They should enable to identify potential problems of sensor calibration or drift and thus to establish the periodicity of maintenance procedures during the FOT. They should also permit to validate the data collection procedure from data acquisition, data transmission to data storage.

   The technical teams involved in the FOT should be in charge of these field tests.

2. The second level of preliminary field test deals mainly with the issue of *assessing the usability and usage of the* systems *under study and of identifying the main critical issues associated with their use in real driving* situations. This is particularly relevant for:

   - Structuring the familiarisation phase of the drivers before their participation to the FOT;
   - Contributing to the design of the questionnaires for the subjective assessment of the systems;
   - Testing and/or improving the various tools developed for data processing, such as automatic identification of critical “use cases” and “scenarios” and video based identification of triggering events or categorisation of road and traffic contexts.
   - Identifying a number of critical scenarios when using the systems, scenarios that could be investigated more extensively when the data gathered from the FOT are processed and analysed.

   This test requires the participation of a sufficient number of drivers (depending on the target population in the FOT) and should be performed in real driving situations. An experimental journey on the road could be designed for that purpose (depending on the hypotheses formulated). This level of analysis provides useful data for

designing the relevant tools for the evaluation process as mentioned above, for estimating the time required for data processing and data analysis and thus calibrating these phases in the FOT. It may be seen also as an opportunity for training the team (s) in charge of data processing. Finally, it represents an important step for testing some of the hypotheses formulated in the FOT and/or for refining them. In this phase, it's important to underline that the drivers used in the pilot study will not be part of the final sample and therefore most of them do not need to be naive.

Psychologists, ergonomists, and human factors experts should perform these tests in close cooperation with the team in charge of statistical analysis as well as the team in charge of developing data processing tools.

To test whether what is asked from participants is realistic, it is a good idea to pilot yourself before letting 'real' participants undergo the testing. Let someone (or several persons) from the project team drive in an FOT vehicle, answer the questionnaires, fill in the travel diaries, etc. This is especially relevant for people working on subjective data collection and analysis.

3. The third level consists of *testing the feasibility of the overall evaluation process* from the selection of the participants through to data collection and evaluation. It is a kind of final rehearsal before the deployment of the FOT. It enables in particular a check of the communication process between the various teams involved in the practical deployment of the FOT of the robustness of the technical tools designed for data collection and transmission and of the robustness of the evaluation tools used in the assessment.

The result of the pilot can be a no-go if too many problems are still present. In this case it could be reasonable to delay the start of the data collection phase and to repeat some earlier steps. This means that there are feedback loops in the piloting process.

6.5 Controlled testing

As described in section 6.2, a power analysis is required to determine the necessary sample size for conducting an FOT. The estimated or simulated frequency of events and the penetration rate are a key element in this calculation. It might prove that a naturalistic FOT is not feasible, due to the low frequency of events resulting in a very high number of needed vehicles or a very long experimental period.

In such cases, one possible option is to allow controlled or semi-controlled testing. This means, that all or a certain group of the drivers are instructed before or during the test execution to behave in a certain manner. For instance a professional driver might be instructed to simulate a car breakdown to trigger the car breakdown warning function in passing (uncontrolled) vehicles. In the controlled approach, the test drivers are called into the test and they are asked to drive the test route with some arrangements. Preferably, the tests will be conducted in real traffic. Some tests, however, must probably be organised on a closed test track. One test may include several runs of the route. Several situational

variables can be fixed in advance. The tests can be designed so that some variables are systematically controlled during the data collection. Based on the practical constraints different levels of control, from totally naturalistic to totally controlled, can be chosen, taking into account that controlled testing breaks with the principle of un-interfered experiments and should be chosen, only if the FOT boundary conditions and/or the power analysis do not allow a naturalistic test of the function under test. Controlled testing can also be used as a supplement to naturalistic FOTs.

Table 6-1 provides an overview of differences between controlled tests and naturalistic driving studies [DRIVE C2X DOW]

Table 6-1: Complementary uses of naturalistic and controlled tests in cooperative system evaluation

Comparison criteria	Naturalistic studies	Controlled tests
General versus experimental design	Normal day-to-day driving. Data collected continuously. Usually same drivers for long time (or full study). Do not bias drivers but get there natural response and acceptance. In cooperative setting assumes vehicle interaction often enough for statistical significance.	Controlling the exact studied scenarios and interaction between vehicles. Easier to have different user groups (old, young etc.) but for shorter times. Necessary for function and technical evaluation. Controlled number of tests gives easier statistical control
Acceptance - long term	Yes, but care should be taken not to influence by administering many questionnaires throughout study.	Not possible. Short study time.
Acceptance - short term	Possible, but takes calendar time since interactions cannot be forced.	Yes, but care should be taken to limit too many repeated interactions over short time. May give unpredictable results.
Impact on environment	Yes, as long as enough system interactions happens. Compliance to system possible to study.	Yes, but assumes driver compliance.
Impact on safety	Yes. Will likely result in wide variety of situations. Possible to compare treatment/baseline for e.g. Crash relevant events. Naturalistic distraction and TTC	Yes, but difficult to cover crash relevant events, distraction and compliance.

Comparison criteria	Naturalistic studies	Controlled tests
	distributions (including compliance)	
Impact on efficiency	Yes, variety of situations and compliance can be included. Need to have large baseline to compare with	Specific situations may give good statistical base for these situations. That system compliance will be almost 100% has to be considered.
Impact on mobility	Yes but needs to look at long term and assume enough interactions. Questionnaires relevant.	Difficult since compliance is very high and drivers are told to perform tasks. Questionnaires relevant.
Driver behaviour - route choice	Learning effect can be studied. Difficult to know if route choice is an option. Indirect effects.	Difficult since drivers will do what they are told.
HMI	Short and long term HMI usage and acceptance possible. Evaluation if reaction time changes over time. Time on task possible if video.	Short term HMI usage and acceptance. Easier to evaluate in controlled environment but long term difficult. Necessary before naturalistic deployment.
Usage (function)	Are the users turning it off over time? Compliance to system information possible to study.	Difficult
Technical evaluation of wireless communication	Study of robustness possible, but in-depth analysis difficult due to uncontrolled scenarios.	Necessary to study in depth communication aspects.
Function validation	Long term functionality evaluation possible but impractical for optimization and technical debugging. May give optimization parameters long term and give information about function validity for real traffic scenarios.	Necessary to validate and optimize functions to technical boundary conditions. Give fast and reliable results that can be quickly followed up.

6.5.1 Operationalisation of tests

Controlled testing requires a strict operationalisation process from the high-level hypotheses down to the individual tests to be performed. A three-step process is advised:

In controlled tests all drivers are instructed to follow a defined *test scenario*. This scenario is created from the hypothesis defined and tries to provoke a system behaviour, which causes the activation of the function to gather data needed to prove or disprove the hypothesis. The scenario should therefore contain:

- Functions addressed
- Hypothesis addressed
- Description of desired situation
- List of desired participant types
- List of desired vehicle types
- List of vehicle groups (e.g. one group as broken down vehicle, one group for the passing vehicles)

It is sufficient if scenarios are described in non-formal text. However it might be advisable to use a pre-defined scheme to describe them. To follow up in the operationalisation, the scenarios have to be further refined into test scripts. A *test script* builds upon one scenario and maps it to a given area and a given project setup. To generate this test script each group has to be mapped onto the road network in the test site. A route is created, which defines for each group, where the vehicles will exactly drive and what timing is desired/expected.

A baseline can be created by assigning a separate control group to the test script with systems switched off.

This test script therefore contains:

- One route for each participant group with timing information (including individual vehicle timing offset)
- A desired state for the functions to test
- A desired state for the logging and monitoring systems

In a final step, the test script is turned into a *test case* before actually starting the test. For this test case the actual drivers and vehicles are assigned to the groups. Also a date and time for the test case is fixed. One test script might be scheduled several times as a test case to gather enough qualified data to filter out outliers in the execution. Drivers and vehicles may change for different test cases of the same test script. In the fig. 6.2 an example is shown for the process explained in this section.

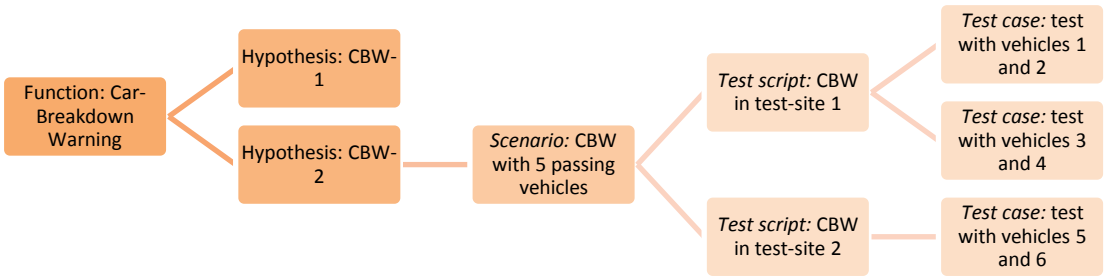

Function: Car-Breakdown Warning
Hypothesis: CBW-1
Hypothesis: CBW-2
Scenario: CBW with 5 passing vehicles
Test script: CBW in test-site 1
Test script: CBW in test-site 2
Test case: test with vehicles 1 and 2
Test case: test with vehicles 3 and 4
Test case: test with vehicles 5 and 6

Figure 6.2 : Operationalisation of test scenarios

6.5.2 Operationalisation tool chain

In larger FOTs a dedicated set of tools is highly advised for the operationalisation process.

In a scenario editor tool all scenarios can be entered in pre-defined fields. These map to the textual information needed to describe the scenario , but also define formal aspects, such as desired number of test iterations or whether a pre-validation in simulation is necessary. It should also list the necessary performance indicators.

The script editor tool is a map-based tool. It loads scenarios and maps the implicit information on what should happen to explicit routes in one specific location. For each of the driving groups one route needs to be created. Intelligent mapping tools allow to use the underlying road network data in the map (e.g. OSM, GMaps) to automatically follow the street. To get a first idea of how the script will perform a real-time minimal simulation can be used to see virtual vehicles move on the defined routes. Thus synchronisation between groups can be reached to successfully create the desired situations.

The script editor tool also creates log profiles to be taken during the test based on the performance indicators contained in the scenarios. For this process the measures needed for all performance indicators are merged. Sophisticated scripts can also contain time-bound or location-bound markers, which are executed in the vehicles once it passes the given point. These markers are used by the test system for instance to trigger:

- a change of log profile (e.g. extended logging, when entering the test area)
- a driver instruction
- activation or deactivation of functions (e.g. for control group)
- a synthetic function behaviour (e.g. turn on the broken-down vehicle warning)

In a final step of operationalisation the test script has to be mapped to the current test site situation. A test case is generated from the test script by allocating available vehicles and drivers to the groups shortly before starting the test. This should not be done in advance, since fluctuations in vehicle pool and drivers are to be expected for larger fleets. The tool chain can support this with a dedicated control connection to the vehicles.

6.5.3 Test execution

In theory a controlled test can run unsupervised. In practice controlled tests need live supervision to have an acceptable success rate. (Note, that a test is determined to be successful, if the desired scenario has been created – not necessarily if the function was triggered)

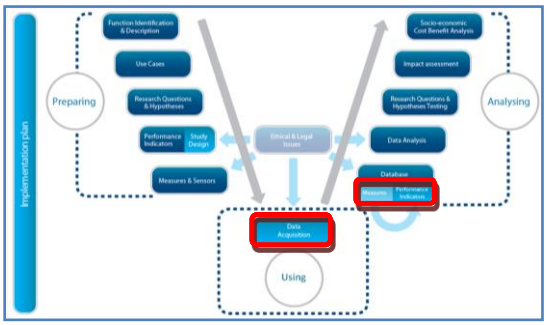
The supervision of a controlled test is preferably managed with a *test control* tool. This tool displays in real-time the status of all participating vehicles (*monitoring*) and the selected test case. Thus, the operator can monitor test progress and determine deviations from the original script.

A way to directly interact with test drivers is desired. Using the same connection as the monitoring data, the test control tool can send messages back to the vehicles. These messages can contain:

- Textual instructions to the drivers (to be displayed on HMI)
- Voice instructions to drivers
- Scenario script and test case information, e.g. Test name, Schedule, Route information
- Trigger information for the test-system itself (log profile changes)
- Trigger information for the system under test

It has to be decided, if driver instructions are necessary for the FOT. If so, they can be either displayed on the system HMI or on a dedicated device.

7. Guidelines for Data Acquisition



This section aims to provide guidelines and recommendations for how to handle data in an FOT[FW] study. Data acquisition, data storage, and data analysis tools will be covered here.

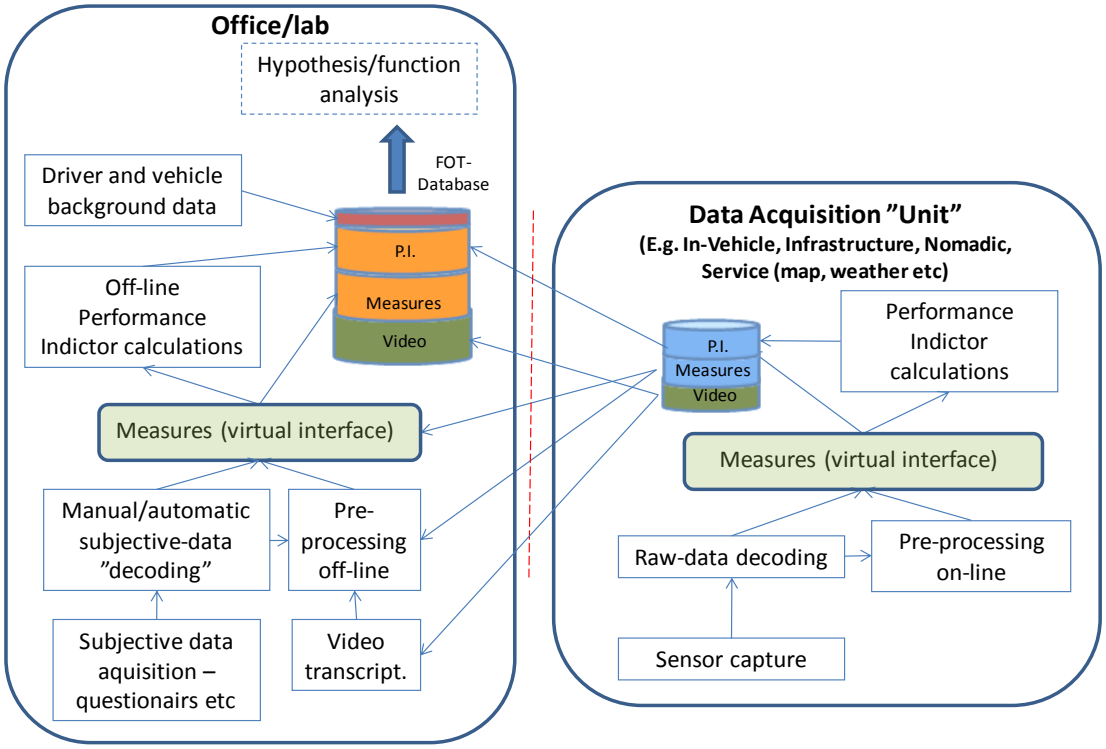


Figure 7.1: Data structuring

Please refer to Figure 7.1 for an overview of a data handling structure for an FOT, and for the naming conventions used in this document. The example data structure above includes data from an electronic data acquisition system, as well as collected subjective data. The Data Acquisition Unit (on the right) comprises sensor systems[FW] requiring raw data decoding. The raw data may then be pre-processed, in this case by low-level data processing such as simple filtering or calculation of directly derived results. Both raw data and pre-processed data (derived from raw data) are stored in the same format locally and can be kept locally, until moved from the Data Acquisition System (DAS) to the main storage, and used for analysis.

Acquisition of subjective data may also be performed. *Subjective data is also considered as acquired from a sensor.* This data can similarly be subject to manual or automatic decoding, database storage (pre-processed or not), and subsequent use in performance indicator[FW] calculations.

It is important to create documentation of developed tools and processes in order to enable their usage after the project end (e.g. usage in follow-up project) and also to provide detailed test protocol with all relevant test scenarios and potential risks.

7.1 The measures and sensors tables

FESTA recommends use of the FESTA matrix, a spread sheet in Excel format containing three tables: “performance indicators”, “Measures”, and “Sensors”, that may in a later stage be utilised to create a relational database. The sensor table should be used in preparation of, during, and after an FOT project, in conjunction with the other two related tables (see Appendix 1 in Deliverable D2.2). Properly handled and thoroughly implemented the tables are valuable tools for data structuring and for data requirements specifications, and for identification of connections between sensors, measures and performance indicators.

The FESTA matrix is intended as an aid and gives examples of commonly used sensors, together with specifications. It is not exhaustive and future users are encouraged to modify, expand, or limit the matrix to suit the particular FOT.

7.2 Data acquisition

Methods of data acquisition in FOTs include methods to collect background data, digitally acquire data from sensors, and subjective data (such as data acquired from questionnaires). In addition, data in the form of manually or automatically transcribed data and reductions of collected data is also considered sensor acquired data (but with a manual sensor – the analyst). In FESTA all the data sources mentioned above are considered sensors. Subsequently can all data be acquired, stored, and processed in a generalised way.

All of these different data types are used to support the hypotheses[FW] defined for the specific FOT. The data to be collected should be defined and based on research questions[FW] and hypotheses.

7.2.1 Background data acquisition

The background data about the driver is crucial and needs to be collected integrated in the driver interaction procedure. Due to privacy issues different parts of the background data may or may not be suitable for storage in a database, or be used in statistical and other forms of analyses.

Data could be gathered by interviews and/or questionnaires, by different tests, or by specific instruments. The driver background information should be considered as acquired from a sensor, and preferably be added into the database and to the sensor matrix, though subject to privacy issues.

7.2.2 In-vehicle data acquisition

An in-vehicle Data Acquisition System (DAS) is needed in FOTs where the focus is to study in-vehicle systems by collecting data from the systems in the vehicle and also from sensors mounted on the vehicle such as video cameras. A suitable DAS can differ from study to study

and a specific solution cannot be recommended for all types of FOTs. See section 3.1.2 in D2.2 for a list of different DAS solutions.

The guidelines and requirements in this document are based on experiences from FOTs using some kind of in-vehicle data acquisition.

7.2.3 Nomadic devices

A nomadic device (ND) or an aftermarket device could be either part of the function/system under test, or it could be part of the data acquisition system, acquiring specific FOT data.

Nomadic devices can also be used as data storage tools as they are easy to install and use on different kind of vehicles. If the vehicle has a dedicated gateway for ND, this option can be used for capture of further vehicle related data.

Using the local wireless connections, the storage capacity of ND could be extended with large capacity hard disks. A possible drawback of a ND, when used as a DAS in itself, is that test subjects must remember to bring the ND to the vehicle every time he/she uses the vehicle.

7.2.4 Subjective data acquisition

As explained before, subjective data are also considered as “sensor” data in the scope of the FOT methodology. All subjective data should therefore be stored and handled logically as if it were collected from a “real” sensor. Subjective data may include data acquired from the test participants in different ways. Results from interviews, questionnaires and focus groups are typical types of subjective data. In evaluations of cooperative systems, consideration should be given to collecting subjective data from other actors such as control room operators.

The result from the subjective data acquisition should preferably be stored in an electronic format. Electronic compilation of the questionnaire may be considered to reduce the overall manual work and cost, maybe by using web based tools.

For subjective data to be stored, the following related information is required:

- Date and time (hh:mm) of test start
- Date and time (hh:mm) of test end
- Subject ID code
- If present, reference to objective data (file name, location)

It is important to think about processes and tools for gathering questionnaires at the beginning of the project

- e.g. provide smart phone apps to fill out questionnaires
- e.g. keep drivers in the analysis loop

7.2.5 Real time observation

In this context, real time observation data is data collected by an observer that directly or indirectly (in real-time or afterwards – for example on video) is observing the drivers and systems to be evaluated. The data acquisition process is usually relatively manual but the results should be transferred to digital format and uploaded to the FOT database for further analysis.

Real time observation data help provide a more detailed picture of a driver's behaviour, as well as verifying the information gathered by other instruments. As the overall purpose of an FOT is to collect information on as natural driving as possible, the real time observation is almost exclusively done through video cameras hidden in the vehicle compartment, as an observer physically in the car will impose the risk of the driver not acting the way he or she would otherwise. Direct real time observations must therefore be carried out with great care and as unobtrusively as possible, or avoided completely.

7.2.6 Additional data sources in Cooperative systems

The cooperative systems architecture implies the possibility of additional data sources. Specifically RSUs connected to proper sensors may provide traffic information and environment data. RSUs connected to Traffic light controller are able to provide traffic light phases and Intelligent traffic control centre dispatch traffic information and alternative routes. It is important that all data records contain a time stamp synchronised to GNSS clock.

7.2.7 Acquisition of infrastructure data and other services

7.2.7.1 General aspects

The infrastructure can be equipped with sensors to detect e.g. traffic or weather conditions. Data from such systems can be collected in raw format or in an aggregated form. If data is collected both on the vehicles and on the infrastructure separately, it is necessary to synchronise the two sets of data. It is recommended that *GPS time* is used as the synchronisation source.

7.2.7.2 Infrastructure

It is in many countries required to contact local road authorities before the installation of equipment close to a road. Working close to or on roads may (depending on country) require special training or licence. In some countries it is even required to use a special company or local road authorities for any installation work close to or on roads.

7.2.7.3 Services

When using such sources it is recommended for traceability (during and after the project ends) to record information about for example version of service, update rates and resolution/precision of the information they have during the duration of the study. It is also recommended to invite the service providers for discussions and possibly partnership in the FOT.

7.3 Specific sensors

7.3.1 In-vehicle video

Most state-of-the-art FOT experts consider video as a very important data source for the identification of driver behaviour and reactions, as well as for the process of analysis to understand the underlying context with regards to the surrounding. When a certain situation or event has been identified for evaluation of a particular system, the video provides the analyst with information about the context of both driver behavioural aspects and the interaction with the environment (if external video is used).

Video can be used in mainly two ways:

- Driver monitoring: Firstly the driver eye/head movements in relation to the vehicle/environment/context, and secondly the driver interaction with the vehicle and other driver actions (pedal, gear shift, steering wheel handling, mobile phone, eating, etc.)
- Environment/contextual monitoring: Helps understanding the driving contexts by collecting information about the surrounding traffic.

For a specific FOT the number of cameras needed, the needed resolution, the views captured by the cameras, etc., should be *defined by what is needed to address the hypotheses*.

To find the correct requirements thorough investigation is needed. Evaluate the results using calculations of predicted storage needs, as well as evaluation of video image quality and set requirements accordingly. In this evaluation experts from both study design/analysis planning as well as the analysis team should be included.

Generally the following parameters affect video quality, and need to be considered: picture resolution, frame rate, colour settings, “regions of interest”, bit rate strategy, bit rate limitations, “quality” strategy settings, and other video compression method related settings.

For more exhaustive methods for defining the video requirements, for information on potential camera related quality problems involving interlacing, and further information on video compression, see section 3.2.1 in Deliverable D2.2.

7.3.1.1 Recommendations

Make a thorough analysis of video acquisition requirements. Set requirements necessary for each individual view, to possibly achieve a first limitation of video data. Choose a well-tested hardware video frame grabber/compression solution, and choose compression suitable for the FOT. The MPEG-4 part 2 and part 10 (H.264) compression feature sets have been used successfully in other FOTs, as well as MJPEG compression.

Video data analysis is normally complex. For this reason, it is recommended to limit video analysis to minimum and try as far as possible to use automated processes.

7.3.2 Internal vehicle bus data

Most modern vehicle manufacturers features one or several internal vehicle networks such as CAN, LIN, MOST or FlexRay. An internal network may carry large amounts of useful information for the FOT. However, there are several concerns with accessing and ascertaining quality on data from an internal vehicle bus.

7.3.2.1 Accessing the vehicle bus

Accessing information from an internal vehicle bus can be highly complex and even void warranty if it is done without the OEM's permission and supervision.

7.3.2.2 OEM cooperation

A description of the entire vehicle network will often contain proprietary information, and may reveal detailed information on specific functions and the vehicle system architecture. Thus, non-disclosure agreements are typically required and can be hard to attain.

An option to using the entire vehicle bus description is that the OEM only provides the description of a selection of signals which enable access to the most important data. Still, however limited the access to the vehicle bus is, there may still be proprietary issues.

By using a bus gateway, the OEM will be able to extract data from the bus without providing any information about how the bus actually works. The gateway is programmed by the OEM to read certain information, decode it and then pass it on to the FOT logger equipment. An illustration of the process is shown in Figure 7.2.

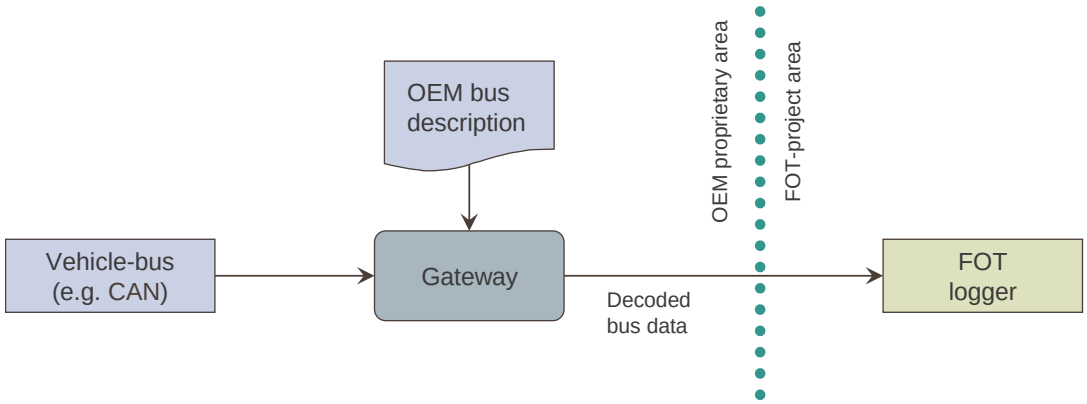


Figure 7.2: Illustration of process for securing proprietary vehicle-bus data using a gateway.

When physically connecting to the vehicle bus, it is important to follow applicable standards of the bus in order to prevent interference that may reduce network functionality (and thus warranty). It should also be stressed that every bus implemented by an OEM might not necessarily fulfil the standards in every detail.

7.3.2.3 Sensor specifications and details

When acquiring sensor data from a vehicle bus, the information is passed through several stages before it can be read from the bus (see section 3.2.2 in D2.2). These stages are likely to affect the signal value both in terms of amplitude and frequency and thus need to be closely observed.

To ascertain that qualitative measures are attained, plenty of contribution is required from the OEM. A thorough description of a vehicle bus and its ECUs will in many cases require involvement by subcontractors as well. A list of required details for successful data acquisition from the frequently used CAN bus is provided in section 3.2.2 in D2.2.

7.3.3 Automatic in-vehicle driver monitoring

7.3.3.1 Head/eye tracker

Two of the main issues for many systems are visual distraction and the effects the system has on the driver attending to traffic versus to the system. By using eye or head trackers in the vehicles, continuous data of some driver state/attention measures can be obtained. The problem with head and eye trackers is mainly the risk of significant data dropouts due to limitations in driver head and gaze tracking.

State-of-the-art head/eye-tracker technology today is relatively expensive, but the benefit of using such a system should not be underestimated. It is recommended that head or eye tracking systems are employed in FOTs where driver state is an issue.

Using an unobtrusive system is a requirement for head/eye-tracking systems for FOTs. The driver should not have to initiate the tracking system or wear any device; the system should be as inconspicuous as possible to the driver. Also, the system should not require any manual calibration in the field.

FOTs using eye-tracking devices have continued to find this a challenge.

7.3.3.2 Other

It is strongly recommended that adding sensors that the driver has to put on and wear should be *avoided* to assure as natural driver behaviour as possible. (For further examples see section 3.2.3 in D2.2.)

7.3.4 Extra analogue/digital data sources

Access to certain information that is not available on vehicle bus systems is often needed. An analogue/digital I/O device for data acquisition is thus required. Also, anti-aliasing issues need to be addressed. The requirements for the different kinds of extra analogue and digital sensing need to be defined in the study design and will not be covered here.

7.3.5 Radar and other non-video environment sensing

7.3.5.1 Sensors already integrated (by OEM/Road Administration)

If a required environment sensor is integrated into the vehicle or the infrastructure, great effort should be spent on trying to add the sensor data to the used data acquisition system. In this integration several issues have to be considered (for details see section 3.2.5 in D2.2):

- OEM allowing/disallowing access to data.
- system interfaces: Some low level sensing information (e.g. object tracks from radar) may require special interfaces to be implemented both in hardware and software;
- system comparability: If the studied vehicles have different OEM integrated systems, they will provide different quality of data as well as different resolution, range, field of view etc.

7.3.5.2 Add-on environment sensing

If additional sensing needs arise, please consider the recommended process for adding sensing:

- Use the sensor matrix for sensing needs and identification;
- Requirements on the extra sensing need must come from hypotheses and performance indicator requirements;
- Additional considerations: Does the extra sensing require additional interfaces? Does it require information from vehicle buses? How can the sensor be integrated without significant effort? Does it require repeated calibration?

General guidelines for specifications are difficult to define. However, for example radar and other object tracking sensing systems, required field of view, radial and angular resolution/precision is important to define based on the hypothesis. For further suggestions, see section 3.2.5 in D2.2 .

7.3.6 GPS

A GPS device can provide a *GPS Time* reference time stamp and a difference between the GPS time and UTC. It is highly recommended that this is used for synchronisation within a data acquisition unit as well as between systems (in-vehicle, ND, infrastructure and services). Information about the present local time zone can be useful for subsequent analyses and synchronisation with non-UTC devices.

Multipath propagation of the GPS signal is a dominant source of error in GPS positioning, since they depend on local reflection geometry near each receiver antenna. In most cases these can be corrected to a large extent with DGPS solutions. The errors depend on time of day, and satellite positioning (in zenith or low orbits) as well as other atmospheric disturbances.

7.3.7 Audio and driver annotation

Continuous audio recording is potentially a significant privacy issue and is not recommended. In state-of-the-art FOTs drivers have had access to a comment button on the dashboard, which, when pressed, would start recording any verbal comments during a pre-set number of seconds (usually around 20-60 seconds). Use of the button could be encouraged if the driver feels that it is warranted or at agreed (critical on non-critical) events. Be sure to inform the drivers consistently about the annotation possibility and provide some simple guidelines on how to offer the comment.

7.3.8 System function/status

A system under evaluation, such as an LDW, ACC or FCW, needs to be continuously monitored to ensure that it is operating properly. The system status signal will thus form a measure to be recorded in the data acquisition system. The status signal will depend on the specific system and needs to be provided by the system manufacturer or vehicle manufacturer, thus requiring strong collaboration with the actual provider.

7.3.9 Vehicle metadata

Information about the studied vehicle is important for analysis and study design. The recommendation is that for each type of study, systems, functions and specifications that may act as confounding parameters in a specific analysis should be stored. During analyses, these parameters may have confounding effects if an inhomogeneous test fleet is used. Examples:

- FOTs with focus on safety: Information should be stored about integrated systems that may contribute to driver distraction,
- FOTs with focus on environmental issues: A more powerful engine, automatic gear shift or four-wheel-drive, are most likely to be confounding parameters in an environmental analysis.
- Experts have also stressed through the seminars, that wider metadata – from weather conditions to prevailing economic conditions is also essential for interpreting data post-study.

7.3.10 Coding/classification/transcription

As part of the data reduction and analysis process, described elsewhere in this handbook, sections of time will often be labelled with classifications according to a coding scheme or syntax. Depending on the study, sections of time can be given categories such as “crash”, “near-crash”, “incident”, “Curve speed warning”, “lane change”, “crash avoidance by steering”, etc. When classifications are made they are often saved and thus become a new data source which is added to the database. For example, an index indicating all instances of lane changes in the dataset can be created and saved. Regardless if the classification of data is performed by a human analyst transcribing video or by an algorithm applying kinematic trigger values to the data, this process of classification should be seen as a type of sensor providing a new data source. Thus, it is comparable to other types of off-line performance indicator calculations (see Figure 7.1). It is recommended to plan that these new data

sources or measures will be created during the data reduction phase after the data has been uploaded to the FOT database.

7.3.11 Geographical Information System (GIS)

One of the lessons learned in state-of-the-art FOTs in the US is that geographical information, such as road curvature, roadside embankments and other on-and off-road information has been underestimated as a valuable source of data in the analysis. Contextual indicators of events and situations identified in the analysis provided added insight into both behavioural aspects and how the infrastructure influences system performance. (See section 3.2.12 in D2.2 for more information.)

GIS derived information about the current road (road type, speed limit, rural/urban, banking, curve radii, etc.) could be used directly in the vehicles as separate measures. By doing this on-line, the necessary post-processing is reduced, but requires additional software and possibly hardware. One advantage of performing this on-line can be that the absolute position (e.g. GPS) does not have to be stored, thus reducing the some privacy concerns. An advantage of using commercial navigation software as part of the on-line map-matching and information extraction can be that there is no need for potential in-project development of map-matching algorithms. This technique has been used in some state-of-the-art FOTs. It is important to validate that map-matching and other GIS data extraction is done in a proper way.

7.3.12 Communication unit

A communication unit is a device, which provides vehicle to vehicle and vehicle to infrastructure connections in cooperative system FOTs. It also has to be considered as a source of measures for all communication-based aspects. Typical measures include received and sent messages, network traffic congestion, number of connected nodes etc.

7.3.13 Application Unit

The Application Unit is a device to run developed functions. Typically this is an x86-based computer in ruggedized configuration. If an application unit is deployed in an FOT it also provides a variety of measures. All developed functions should be providing measures of their state and of key variables used. The underlying Facility layer (e.g. vehicle access, position access, local dynamic map) is also a valuable source of measures.

It has to be decided, if it is feasible to also run the data acquisition system directly on the Application Unit, which simplifies the access to all key variables. On a downside the DAS might consume too much processing power or drive capacity, thus interfering with the functions.

7.4 Mechanical requirements

The following mechanical guidelines and requirements are primarily applicable for FOTs with in-vehicle Data Acquisition systems.

7.4.1 Size and weight

The system should preferably have negligible effect on the driver's use of the vehicle – including limitations in trunk space.

7.4.2 Connectors and interfaces

State-of-the-art FOTs state that from experience as much as 80 % of the DAS hardware problems can be deduced to physical connector issues (to the DAS and to peripherals). It is recommended that connectors with some locking between connector genders are used. Cable pull-relief should be used when possible.

7.4.3 DAS mechanical cover and ease of access

It is recommended that a layman without tools is able to find and have visual access to any indicator LEDs on the DAS. Also, having the possibility to connect interface devices without having to remove covers is preferable.

7.4.4 Crashworthiness and vibration resistance

For all FOTs the minimum requirement for ruggedness is that the entire system should operate under the normal driving conditions for the specific FOT. including the harsher situations of normal driving.

7.4.5 DAS environmental requirements

Environmental requirements for the DAS mainly concerns temperature. If the DAS is placed in a shielded location, the need for water resistance may be negligible, although the DAS internal parts should be able to withstand reasonable levels of condensation. If applicable, it is recommended that a simple dust/particle filter is placed by the main air intake of the DAS. Important considerations include sufficient cooling of the system due to internal heat generation and for ambient temperature. The cooling needed depends on the processor used and the workload. If the FOT is using video compression done in software, the need for processing power may be drastically higher than if only simple signals are recorded. High processor load generates significant levels of heat. The need for forced ventilation is depending on the DAS mounting position.

Too high and too low temperatures (both static and transient) do affect the DAS. Components with moving parts need special attention. Hard drives are some of the most sensitive components. Automotive grade hard drives are available, although somewhat more expensive than normal consumer hard drives, and with limited storage capacity. Flash memory cards or solid state drives (SSD) are available in increasing capacities and are clear alternatives for operating system hard drives and for storage in lower data volume FOTs.

7.5 Electrical requirements

7.5.1 Power management

The main requirement for the DAS setup is that the installation should never affect or impair normal vehicle function. *This requirement should be enforced in all environmental conditions* and is one of the most critical issues for drivers to accept equipment to be installed in the

vehicles. *If* FOT *installations are rumoured to impede on the trustfulness on vehicle operation, few people will volunteer in any subsequent study, and the study at hand as well as subsequent studies may be compromised.*

The entire DAS installation may not draw power so that the vehicle battery charge falls below the level of being able to start the vehicle. Care should be taken so that the system does not draw power (or only minimum power) if the ignition has been turned on but the engine did not start, or if the engine stops without the ignition key being removed. State machine evaluation should be applied.

7.5.2 Interference with in-vehicle equipment

In all in-vehicle installations the aim should be to minimise or manage without the use of AC powered devices. The DC/AC and DC/DC converters are sources of electromagnetic noise and may affect both standard in-vehicle equipment (such as the FM radio), and the FOT installation itself.

Attaching any equipment to the in-vehicle bus systems has to be done very carefully. Transmitting data on vehicle buses should in most cases *not be needed or not done at all* in an FOT implementation. Failure to adhere to this might be dangerous and result in vehicle operational malfunction that may result in significant cost, injury or death, or produce other very unwanted results. Even adding only listening/eavesdropping devices should be approved by the vehicle manufacturer before being implemented.

7.5.3 Laws and regulations

Depending on the FOT study at hand different regulations will apply. In some cases CE certification of the FOT equipment may be necessary, and each project must verify what is applicable to the specific study. If the vehicles are to be driven in non-EU countries the specific regulations for each region should be verified.

For wireless communication and for some sensing systems there are regulatory restrictions on transmitting electromagnetic radiation. A few example of sensing systems that have direct regulatory restrictions are LIDAR and RADAR. The restrictions may be based on electronic interference or harm. This has to be taken into account for each individual FOT sensor setup. Still, for each instrument and jurisdiction, care should be taken to investigate the applicable regulations.

Several countries have regulations for equipment employing radar or laser technology on public roads, as these can affect effectiveness of for example authority speed surveillance instruments.

7.6 DAS data storage

7.6.1 Storage capacity estimation

The main aim of the *storage capacity estimation* is to guarantee the availability of free space for recording the vehicle data. Storage capacity depends on the following factors: number of recorded signals, sample rates of the recorded signals, sample size of the recorded signals,

data collection method, driving hours, data size reduction (filtering/compression), and data deletion procedure.

Ideally, the sample rate for each signal should be the lowest possible able to guarantee no information, relevant for answering the FOT hypotheses, is lost in the sampling process. Some sample values are reported in the FESTA Deliverable D2.1 for CAN and sensor data.

Using a dynamic sample rate and a high sample-rate buffer would make possible adapting the resolution of the recorded data depending on the event[FW]. The drawbacks of using dynamic sample rate are that the recording system complexity (and probability of faults) increases, more post-processing on the data will be necessary to handle the different sample rates. Database design and search becomes more complicated.

The data collection method can be continuous, or limited to specific events of interest, for example time intervals in which the lateral acceleration is above a certain threshold or in which the vehicles enters a curve with speed above a certain threshold (see section 3.5.1 in D2.2). In FOTs where only triggered storage is used and where the driver subjectively triggers in some way, it is important that past data also is recorded by using for example a ring buffer (buffering one to five minutes in the past). The level of pre-trig time will differ between projects, and should be defined based on the hypotheses for each study.

Note that events such as occurrences of safety systems warnings can be considered. The probability of such events should be part of the knowledge of the FOT designer and should be used for the estimation of the storage capacity.

Driving hours depend on the nature of the driver and the vehicle. See section 3.5.1 in D2.2 for examples.

Compression algorithms can help to reduce the data size. A lossless compression (such as zip) can be used for CAN and sensor data whereas a lossy compression (such as MPEG-4 and MP3) is normally acceptable for voice and video, respectively. The drawback of compression is that the complexity of the system increases and new possible sources of error and malfunctioning are introduced.

A safe data deletion procedure implies that no data is deleted in the vehicle until a copy of the data has been backed-up, verified, and stored in a safe place.

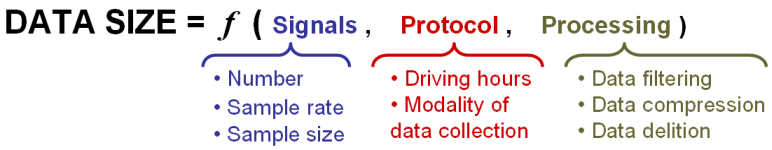

DATA SIZE = f (Signals , Protocol , Processing)
• Number
• Sample rate
• Sample size
• Driving hours
• Modality of data collection
• Data filtering
• Data compression
• Data delition

Figure 7.3: Factors influencing data size

Please refer to section 3.5.1 in D2.2 for storage capacity estimation equations, as well as data size estimation examples.

Storage capacity may be depleted and the intervals for retrieval and uploading may present some variability. These factors should be taken into account by guaranteeing enough tolerance on the final storage size. Since no space to record data would result in data loss a 20 % to 50 % on storage size tolerance is recommended.

In some studies where the levels of allowed data loss are very small it is recommended to use direct on vehicle data backups. This can be done for example by using several storage media with a data mirroring solution (e.g. RAID).

When triggered data acquisition has been chosen as the data collection method, great care has to be taken to define the trigger definitions. Even if triggered logging is used for the evaluation of effects, most studies will require baseline data. It should be possible configure and acquire baseline data, using the same DAS, also for these cases.

FOT activities that use any type of triggered data acquisition and have high data-rate data sources will need significant amounts of main memory to handle the necessary ring-buffer for pre-triggering storage. Moreover, estimations of needed storage space and data retrieval/upload frequency are affected (See section 3.5.2 in D2.2.)

7.6.2 Data retrieval/uploading procedure

Data retrieval/uploading procedures are needed to make sure that all collected data is backed-up and stored in a safe place in order to minimise data loss. The aim is to prevent data loss, verify data completeness, and to prevent data storage waste (caused by double storage).

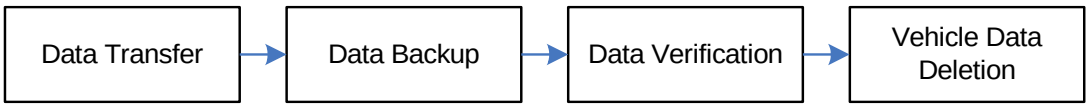


Figure 7.4: Data retrieval /uploading steps

7.6.2.1 Data transfer

Data transfer is aimed to assure that a copy of the data collected in the vehicle is stored in a safe location. *Data back-up* is aimed to prevent data loss by having a multiple set of the data stored in different safe places. *Data verification* is aimed to assure that no data was lost during *data transfer* and *data back-up*. *Vehicle data deletion* is aimed to ensure that storage space is newly available in the vehicle once the data has been safely transferred and backed-up.

Depending on the support used to record the data in the vehicle and the data size, different data transfer modes can be implemented. For a list of potential modes, transfer rates, and ranges please see section 3.5.4 in D2.2.

Generally data transfer poses two main problems: It may be time-consuming and data can be lost during the transfer. The following table presents an assessment of different transfer modes.

Table 7-1: Pros and cons for different data retrieval/uploading modes

Data transfer mode	Time efficiency	Cost efficiency	Technical complexity	Data security	Driver comfort
Data is "picked up".	+ Very fast from a vehicle point of view.	- May require to pay someone to pick up the data. - May require an extra vehicle chasing the FOT vehicles.	+ Very reliable because simple.	- Data may be misused between the "pick-up" and the final storing.	- May require the driver to go somewhere or to move the data.
Data is transmitted with wired transmission.	- May be slow for big amount of data.	- May require to pay someone to do the downloading.	- The equipment needs to be safely accessed.	+ Can be very secure if the driver is not able to access the data.	- May require the driver to go somewhere or to move the data.
Data is transmitted with cellular transmission.	- May be slow also for relatively small amount of data.	+ Download can be automatic. - May be expensive depending on the network providers.	- Automatic wireless download is very complex.	+ Can be very secure.	+ The driver does not need to do anything.
Data is transmitted with WiFi transmission	+ Relatively fast - Requires vehicle to be in Wi-Fi AP range	+ Download can be automatic. - Requires Access Point installation.	- Automatic wireless download is very complex.	+ Can be very secure.	+ The driver does not need to do anything.

7.6.2.2 Data back-up

Data should be backed-up and stored in a safe place as soon as it is available. Ideally, the backed-up data and the main copy of the data should be in two different safe places. In some studies where the levels of allowed data loss are very small, it is recommended to use direct on vehicle data backups. This can for example be implemented by using several storage medias with a data mirroring solution (e.g. RAID).

7.6.2.3 Data verification

Due to the potentially huge amounts of data handled, data verification is important since the probability of errors during the copying process is high.

7.6.2.4 Vehicle data deletion

Data from the vehicle should be deleted *only* once the data has been backed-up and verified.

7.6.2.5 Data loss

Experience from previous FOTs tells that data loss at the retrieval/upload stage is common even if it could be almost totally avoided with a robust and well-tested procedure. To prevent data loss during the data upload/retrieval procedure it is important to verify that data is consistent before deleting it from the vehicle. In case data is picked up and the data is not consistent, the vehicle data logger should be checked as soon as possible. Monitor the state of the data logger so that any issues can be recognised and fixed as soon as possible.

7.7 System configuration

Although DAS system configuration needs can be different, one basic requirement that apply to all FOTs is that it should be possible to find and configure a specific DAS after the study has been finished.

7.7.1 DAS inventory management

A system for basic inventory management is recommended for FOTs with more than a few vehicles in use. For such a system to be efficient, sensors, DAS units, vehicles and all other equipment need to be included, as well as relevant supporting procedures developed. For any one point in time it should be possible to deduce the exact hardware and software configuration of a particular installation.

7.7.2 Configuration tools and traceability

In addition to an inventory management system, it is appropriate to employ a system (and supporting management procedures) for configuration management. Software versions and additional information such as (but not limited to) software configuration files, start-up scripts, device calibration files, and other file based data needed for proper operation of the DAS should be stored together with information about the present DAS hardware configuration.

Similarly to above, it is important that traceability of software configuration changes is maintained. In this way, the exact software configuration of a particular installation at any particular time can be reproduced and reviewed.

7.7.3 Switching between configurations

For some FOTs switching between configurations in each vehicle may be necessary. In addition to having to change DriverID when a new subject is introduced to a vehicle there are a few other situations where it may be important to know who was driving or what was done with the vehicle. If this is not separated from the subjects, these trips may be classified erroneously. Examples of situations to consider are: the vehicle driven back for overhaul/maintenance, validation testing just prior to vehicle delivery, and other engineering testing.

If possible, the use of remote desktop tools over wireless (e.g. WLAN or 3G) or wired networking (Ethernet) is recommended for remote administration.

7.8 Acquisition of data

When controlling the power supply to the DAS the start-up and shutdown speeds must be optimised to reduce loss of data. Loss of data can occur both during hardware initiation when no software is started and during hardware termination when no software is able to trigger on a vehicle restart.

7.8.1 Start-up

Normally the data acquisition will start as the vehicle ignition is turned on. In order to minimise the data lost during the start-up procedure, the hardware and software must load and initiate as quick as possible. The start-up time (or the duration where data is lost) should be well monitored and documented, preferably as a property associated with each recorded trip (since it might differ with temperature etc.). Start-up of the DAS hardware shall not be done if the voltage is too low.

7.8.2 Acquisition of data

The DAS hardware should be kept powered on and running during the entire trip. To ensure that the host power system is not overexerted, the power management unit must continuously monitor the power supply and initiate shutdown if a permanent voltage fall is detected. The system must not shut down on temporary variations such as the drop during engine crank. For such circumstances, an energy reserve such as a battery may be required.

7.8.3 Shutdown

When the DAS recording has stopped, the DAS should be kept running for a short time (typically a few minutes) in case the vehicle is started again. Otherwise the DAS hardware should shut down itself as fast as possible. The power management unit should then, after a short interval, cut the power supply to the DAS, regardless of it is properly shut down or not.

7.9 Synchronisation

FOTs are considered to be used for logging of data and not as part of a hardware-in-the-loop component (for example prototype system development). This means that data from the different data sources do not necessarily have to be available for storage close to the real-world event, as long as they are individually time stamped for off-line recreation of the time-line.

7.9.1 Time stamping versus real world event

Latency is the time between when the actual real world event takes place and when the data from each respective sensor is time stamped in the logger. In most FOTs there is a need to specify the requirements for allowed levels of latency, based on the hypothesis. It is recommended to explicitly evaluate what the latency is for each data source (and for some sources individual measures) and compare this with the defined requirements (based on hypothesis; needs to be done as part of the hypothesis and performance indicator

generation). If latency has been measured properly and there is limited jitter, the time stamp time can be corrected by offsetting with the latency.

Note that for data sources that are not controlled by the DAS implementers, such as vehicle CAN, it may be even more difficult to obtain the necessary information for the latency from real world event until time stamping.

7.9.2 Integrated sensing synchronisation

Depending on the methods of analysis and the implementation in the database, the needed level of synchronisation as well as the importance of measuring latency between different integrated data sources in a vehicle will differ. In Table 7-2, issues and methods for calculating the latency for some in-vehicle data sources are shown.

Table 7-2: Methods and issues in calculating latency for in-vehicle data sources.

Data source	Methods and issues in calculating latency
Video	May produce significant jitter/fluctuation on itself and other data sources. Approximate latency can be ascertained using a synchronised LED-light (measured by digital I/O). Preferably hardware synchronised cameras should be used.
GPS	The latency can be calculated very precise since the GPS time is the actual acquisition time
CAN	Difficult to establish latencies for internal systems, but with reference sensors it is possible to get the latency for some measures.
Acceleration/ Yaw rate	Latency can easily be measured using a reference sensor.
Radar	A lab setup with reference sensors of tracked object motion can be used

7.9.3 Synchronisation with nomadic devices

The recommended method to realise synchronisation between the different sources of data is to use GPS UTC time. Thus, all acquired data should be associated with a time stamp that is represented as absolute GPS derived time. It is recommended to use nomadic devices that have easy interface with GPS, so that the absolute time information is available.

7.9.4 Synchronisation of infrastructure systems

Also in the case of infrastructure, the UTC time derived from GPS if useful. Another possibility is to use an on-line time synchronisation service like NTP/SNTP. In the case of a traffic data and safety related FOT, data may have to be stored as raw data, accurately synchronised in time, to allow the reconstruction of the scenario in the following data analysis phases.

7.9.5 Synchronisation of cooperative systems

- Synchronisation of systems with communication between vehicles can also be realised without a central infrastructure. Also here, it is recommended to use the common reference time provided by GPS.

For central ITS stations involved in cooperative systems FOTs and additional external sources it is advised to use a network time synchronisation protocol, such as NTP.

7.9.6 Synchronisation with interviews and other subjective sensors

In many cases it is enough that the interviewers write date and time (hours and minutes) of the interview or questionnaire. If the subject is requested to indicate and/or comment events during the driving (for example by pushing a button), this should be time stamped when logged if possible, to enable synchronisation of the event/comment with other data. The accuracy needed is in most cases less than 5 seconds. For post-hoc structured comments or questionnaires on video or events, it is important to define a process of linking the events to absolute time.

7.10 DAS status and malfunction management

To simplify laymen feedback on system status to the responsible technicians at times of system problems, LEDs or similarly externally viewable information about the system status can help and are recommended.

7.10.1 System status uploads

In order for the people responsible for DAS and sensing in the project to be able to continuously monitor the status of the test vehicles while on the road, a remote (wireless) transfer of the system and sensing status is preferred. (See section 7.14.1 for details.)

7.10.2 Malfunction management

A process for identification of problems in the vehicle, contacting the driver and exchanging sub-systems or the entire vehicle should be developed.

7.10.3 Spare system management

For quick and efficient problem solving it is recommended to keep spare parts for all components pertaining to the DAS. It has been suggested that a 10 % extra number of spare parts should be kept within the FOT, and that this number be added specifically when estimating the total project cost. All spare parts are to be managed by the inventory management system. Testing and calibration of spare parts should preferably be planned and performed within the process. Supporting management procedures need to be developed as well. At least one fully equipped DAS system should be kept “on the shelf”; prepared and calibrated for immediate use.

7.11 System installation

7.11.1 Installation procedures

Before initiating the installation procedures, an installation specification document shall be prepared. The installation specification must in detail describe how each component of the system shall be installed. Specifically, the installation specification shall provide solutions to the following topics:

- Mounting: positioning, means of attachment, accessibility, safety and security;
- Cabling: dimensions, shielding, drawing, mounting, tolerance and labelling;
- Connectors: soldering/pressing, robustness, impedance and labelling to avoid mix up;
- Power supply: consumption, fuse, voltage, source and switching;
- Environmental endurance: effects on electromagnetic disturbances (EMC), temperature, humidity, vibration, shock, electric safety and dirt.

It is of great importance that the FOT system installation is adapted to the requirements set by all other systems in the vehicle. If this is not done properly the installed system could generate disturbances that might void warranty of the original systems – or even an entire vehicle. All systems that possibly may be in conflict with the installed system need to be identified. An adaptation plan must then be developed for each system to ensure that they will be able to operate properly after installation.

The actual installation work needs to be done by operators that are authorised to work on the actual host system, and during the installation work, all changes to the host system (if any) must be documented in detail. Depending on the study and region, the authorising body could for example be the OEMs, insurance companies (voiding warranty etc.), the project and/or a legal entity.

If several vehicles are to be installed, it is recommended to select one vehicle as prototype. The prototype installation will then revise the installation specification continuously during the work.

7.11.2 Installation verification and calibration

When the system is installed it needs to be verified and calibrated before the data acquisition starts. The verification will refer to the installation specification and verify that all requirements are met. Monitoring of all potential sources of interference so that no conflicts are caused is important.

To ensure data validity and quality a calibration and verification scheme is recommended. For data quality aspects it is important that all installed systems of the same category are calibrated and verified using the same procedures. During the verification process a full dataset should be recorded for the analysts and quality management team, in order for them to verify that the installation adheres to the analysis requirements.

7.11.3 Dismounting the system

When the data acquisition is finished and the system is to be uninstalled, the installation documentation shall be used to ascertain that the host system is restored to its former condition. Finally, all proprietary data need to be removed from the FOT system before it is disposed or reused in future projects. Remember to include dismounting costs in the FOT budget.

7.12 Proprietary data

The concerns regarding proprietary data are to keep the CAN/LIN/MOST specifications OEM-confidential and to hide the actual system design to prohibit reverse engineering based on data collected within an FOT project. Regarding the first issue there are two cases to be distinguished.

- When the OEM is strongly involved in the data acquisition process during the FOT execution, the confidentiality of the CAN/LIN/MOST specification is not an issue within the project.
- When the OEM does not handle the data collection by himself, the usage of CAN gateways is proposed. The CAN gateway has to be programmed by the OEM to provide the data from the CAN/LIN/MOST bus according to the agreed logger specification.

The second issue – reverse engineering of functions and systems – is also an issue within the FOT project. Each project will have to handle this and define what is needed. In some cases it may be necessary for the OEMs to handle detailed low level data and aggregate it on a certain level before it is provided to the project partners responsible for data analysis. A general recommendation to future FOT projects is to define in advance, *what level of system data is needed to answer a specific* research question *and whether the involved* OEM*s are able to provide this data to the project*.

In some FOTs OEMs might be interested in the acquisition of additional data, which is not directly related to the project and proprietary to the OEM. This should be allowed. The OEM could separate the additional data from the project data before the data is provided to the further project for analysis.

7.13 Personal integrity and privacy issues in data acquisition and analysis

Recommendations for the definition of necessary legal arrangements depending on the specific FOT are not covered here. See Chapter 3 for further information.

Different levels of data security should be implemented in order to cover personal and privacy issues properly. The data access right of a project partner should depend on his specific role in the project.

Several different levels of data security should be implemented. Driver data allowing direct conclusions about the identity of the driver is considered to have highest requirements

regarding data security. Video data of the driver's face and GPS data are typical examples of the data that belongs to this category. Some types of metadata like the car serial number also belong to this category of data. For this kind of data anonymisation via monikers is required before upload into a database.

It might also be necessary to implement a GPS data based control, which deactivates the video recording, when required (e.g. when driving in countries with specific legal requirements).

7.14 On-line quality management procedures

In all FOTs assuring data quality in the data collection and data management process is very important. The procedures for data quality assurance before, during and after the data collection should be well defined. Specifications and plans should be written for each individual FOT.

It is recommended that a quality management team is appointed for each individual FOT with roles such as: daily quality overview, OEM contact person, subject contact person, DAS and sensor maintenance person, and vehicle maintenance person.

7.14.1 Remote automatic upload

In most state-of-the-art FOT wireless transfer of vehicle and data status has been used, in order to assess the status of vehicle, DAS and data without having to physically access the vehicle when the vehicles in the study are on the road. Different transfer techniques such as simple text messages (SMS) or GSM/3G, have been used for status uploads. A maximum delay from the time of actually collecting the data until it has been analysed for quality and status should be defined. Otherwise the project risks that the vehicles on the field are potentially not collecting the required data. The maximum delay should be set based on the accepted levels of data loss and the length of the study.

For a thorough listing of example variables/measures that may be of interest to store in the vehicle DAS summary files for per trip data upload, see section 4.1.1 in D2.2.

When the data has been uploaded and put into a database trip statistics can be calculated per vehicle or driver. It is recommended to use this to identify extreme/abnormal driver/usage behaviour early in the study so that if necessary drivers can be exchanged. This driver monitoring early in the study is highly recommended so that the study schema of the specific FOT is kept.

If an FOT is to be executed across country boarders and include roaming for the wireless services, investigation of the cost/benefit of using the quality data upload systems outside of the "home country", should be made.

7.14.2 Automatic and manual quality checks

It may be tempting to do the quality assessment fully automatic, but state-of-the-art FOTs have indicated that by doing this there is the risk to contact the driver in cases where the error or anomaly is not significant for the study. Due to this, it is recommended that the

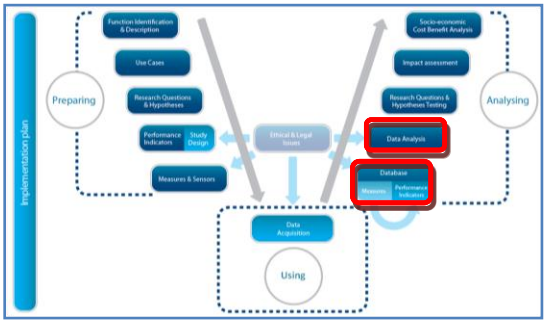
quality assessment should be set up in different steps and that before (if) employing a fully automatic system, the algorithms for the assessment should be thoroughly validated. An automatic thresholds based warning system can be applied for some hard and very important measures. A tool for maintaining the warning system should preferably be checked by a responsible person each day.

For a description of a process for on-line data quality checking, see section 4.1.2 in D2.2.

7.14.3 DAS and sensor maintenance

It is recommended that a specific subject/driver only has one contact person throughout the study. The process for contacting the driver should be clear and preferably there should be a list available within the project with contact information for the contact-responsible for each vehicle. All communication with the driver should then be at least initiated via this person. For a description of a process for fixing DAS problems/maintenance, see section 4.1.3 in Deliverable D2.2.

8. Guidelines for - Databases and Analysis Tools



As state-of-the-art FOTs[FW] has proven that various types of studies demand different data models and hardware specifications, this section will not describe a generic solution for all types of FOTs.

The following sections will focus on database for large FOTs (where thousands of hours of raw data are collected even using different server locations). Such a considerable amount of data, especially if video data is also collected, will test hardware to the limit. A smaller FOT might still use the guidelines in this chapter and apply them to a less complex database.

Recommendations for data handling, especially in connection to data sharing, can be found in FOT-Net 2 D3.2.

When designing data analysis it is important to consider the following general recommendations:

- Create documentation of developed tools and processes in order to enable usage after project end (e.g. usage in follow-up project).
- Provide detailed test protocol with all relevant test scenarios and potential risks.
- Make sure that complete data analysis process (data processing and analysis tools) is developed before piloting phase, in order to test and assess all data analysis step within piloting phase.

8.1 Database design and implementation

8.1.1 Preferred database models

The main challenge of an FOT database is to make thousands of driving hours manageable from a storage perspective and available for ad-hoc analysis. FOTs often have very specific demands, making it difficult to recommend one generic database model. However, the concept of storing the time history data in a single table has been found to be more efficient from the perspectives of analysis and database performance. There are of course need for meta data, user added data (as manual annotations) and events[FW] that give value to the raw data, but the time history data is really the part that put high demands on the software and hardware.

Strategy: A measure equals a column in a table in the database. In order to avoid costly join operations when performing analysis it is also to be considered to keep the database as de-normalized as storage allows. Some data in trips table, such as driver ID or vehicle ID, can be also included in the time history tables. From a storage or relational database perspective*, this is not at all preferred, but when doing analysis it will reduce complexity of SQL queries and also save computing time.*

8.1.2 Data filling in the database

Beside the sensor and video data collected by the DAS, other data needs to be transcribed in the database in order to prepare the database for data analysis.

The documentation of the data shall include both descriptions and format of the data, as well as descriptions of different meta data. This is especially important if the data is planned to be shared and re-used after the project, see FOT-Net 2 D3.2.

8.1.2.1 Trip ID and time data transcription into the database

Trip ID and *time* are cornerstone indices in the database designs. This means that sensor data must be time stamped when inserted in the database. State-of-the-art FOT relational databases use a common sample rate to ensure the validity of trip ID and time. If different sample rates are needed within the same FOT database, the different datasets should be organised into different tables. Furthermore, data with frequency differing from the default one (e.g. 10 Hz) should be clearly marked as potentially incompatible with the main data. If the need to join this data with the main data arises, it is suggested that data from the deviant table is extracted, resampled and inserted into a table with a common sample rate.

8.1.2.2 Additional data

Additional data can be collected, especially if other actors such as infrastructure have incidence in the hypothesis[FW] and performance indicators[FW] defined (see Chapters 4 and 5). This data (e.g., traffic status or road event information in the case of cooperative systems) is usually provided by external entities (e.g., Traffic Management Centres). Collection of this data is necessary to address the FOT needs. In this case, it is highly important to synchronise this data with the one collected in the trip (e.g., GPS time stamp). Data to be collected and its specific requirements should be defined according to the Chapters 4 and 5.

8.1.2.3 Transitional data transcription into the database

Transitional data can be stored separately into tables that only contain data when transitions occur. Despite the potentially huge storage overhead, the trend is to handle transition data the same way as measure data to simplify analysis. Events *data* can be described as shortcuts or pointers to specific events within the database. It is up to each individual FOT to define what an event is and the algorithm that defines it. *Manual annotations* are another way to create pointers to events in the database.

8.1.2.4 Background data transcription into the database

There are two types of *background data* which should be stored in the database model: 1) the driver's and 2) the vehicle background data. Driver data should be stored in the *driver* table but any data to identify the driver should be kept securely and separately (see Chapter 3).

8.1.2.5 Subjective data transcription into the database

To reduce errors, automatic transcription of subjective data is always preferable. Transcription of audio/voice messages is recommended.

8.1.2.6 Time history data transcription into the database

Most of the data in a typical FOT are stored in time history tables. For the *database model 1* (see description above) it is important not to create tables with too many columns. When using the *database model 2* (see description above), specific database tools and functions[FW] could be considered. Examples of these database tools and functions are: table partitioning, block compression, index pre-definition.

8.1.2.7 Events/classifications transcription in the database

An FOT database can consist entirely of events if a triggered data collection approach is adopted (as opposed to continuous data collection; see Chapter 6). In other FOTs, where data collection is continuous, the ability to find and classify events of interest is of central importance. Classification and use of "events" (classified time periods) is an important aspect of FOT analyses (see Section 7.3.10). Some events are straightforward and simple to identify, for example hard braking defined as peak deceleration > 0.7 g, and may not need to be saved as a discrete or transition variable. However, many events involve a considerable amount of effort to find and validate and are worth saving into a discrete variable database or index to facilitate data query and analysis. Event pointers should be saved to speed up analyses and can be used in combination to describe more complex situations[FW] with multiple events (see Chapter 6).

8.1.2.8 Adding tables to the database

Tables are initially created manually based on information in the sensor matrix (see Chapter 5). Any change to the database design must be documented thoroughly.

8.1.3 User data spaces and data sharing

It is vital to keep track of where data is created and manipulated. A copy of the collected raw data should be kept as original data in a read only space to prevent accidental data loss. In addition, it might only be the person responsible for data uploading who can insert new data and the FOT database owner who can delete data. As the analysis work begins, there will be the need for the analyst to store new data (coming from combination or processing of the raw data; see Chapter 6) in a private user space. If this new data is also relevant for other users of the database, a solution to share this data in a project internal space should be implemented. The approval process to share data should be described and meta description of the data is needed (as data origin and function/method/algorithm applied to the data). Some of the data could be of public interest , in some cases on an aggregated level and therefore exported or accessible via web interface. Other data, such as video and GPS, could be made available to researchers using more rigorous procedures. Sharing this data must be approved by all stakeholders in the FOT preferably early on in the project. More detailed information on data sharing can be found in FOT-Net 2 D3.2.

8.1.4 Hardware and storage

It is important to consider that the database server will rest inactive during data collection and then run at 100 % of its capacity during analysis. If the supporting organisation can provide flexible solutions, such as server virtualisation and/or clustering, the FOT study

access to the database when running analysis on the data can be prioritised. When the project ends, machine usage can be set to a minimum until a subsequent study needs to use the data.

Very fast and reliable disks can be used even with a limited budget. In most cases storage at some kind of disk cabinet, NAS (Network Attached Storage) or SAN (Storage Area Network) is most appropriate. A storage setup with some kind RAID configuration should be considered, in order to be better prepared if a disk crashes or some data blocks are corrupted. The database should use faster disks than the file server and using disk cache is recommended to increase the performance of the system.

8.1.5 Risk management

An FOT study can generate huge amounts of data; especially when video is used, and the management must decide on the need for backup and acceptable downtime for recovery of the FOT database. It is up to the steering committee of the study to have a documented backup policy and crash recovery strategy. Further, the backup strategy might need to vary during the lifecycle of the study (collecting phase, analysis phase). If so, each phase and strategy should be documented. Disaster recovery (when local database and backup hardware are destroyed) strategy must be taken care of and there should be an offsite backup of the data.

8.2 Database and data storage implementation

8.2.1.1 Database

Storage of all data but video should be stored in a relational database.

This implementation must consider what to do with data loss from a sensor. Various strategies can be employed: if a sensor gives no data, a NULL value can be inserted. State-of-the-art FOTs suggest that using the last known sensor value makes analysis easier. The problem with data that is actually not valid has to be dealt with (see Chapter 6).

8.2.1.2 Video data storage

A common way of storing video data in FOT-context is to store the video files on a file server and store the links to the video files in the database (to link the videos to the trips).

For very large FOTs, large amounts of video data can exceed the limits of file systems or storage appliances. This can cause extra complexity, for example to add logic (scripts) to enable a single mount point that is preferable from a data management point-of-view. Different file systems and appliances should be evaluated; for instance ZFS or equivalent for almost unlimited shares.

It is worth to examine different video codecs; using an optimal codec can reduce the storage need significantly. The cost (mainly CPU) for re-processing is very high and a large FOT might have need for High Performance Cluster to complete in acceptable time frames.

8.2.1.3 FOTs Hardware

It is recommended to separate the database and video file server in order to configure the hardware individually. Outsourcing system operations is possible; however, the costs for network bandwidth, backup, and administration can be very high.

8.2.1.4 Distributed system at various locations

It is strongly recommended that the database is not distributed. For the database, use a single common database. For video storage, also other options can be considered (see section 5.2.4 in D2.2).

However, due to the location and size of the FOT, it might be necessary to establish a distributed solution for data storage. This is especially true when deploying an FOT in different countries with different local data responsible. In this case, it is recommended to establish a central data server per FOT in charge of gathering all data from all individual databases.

Connection to this central database should be guaranteed so information can be easily transferred from and to this database. A broadband IP connection is then recommended, with simultaneous access.

Synchronisation between the central database and individual databases is to be considered, with automatic synchronisation every certain time period. Manual synchronisation is also allowed, but as a complement to automatic synchronisation.

8.2.1.5 Physical access

Physical access as well as the approval process for access to the hardware must be documented.

8.2.1.6 Logical access

Logical access as well as approval for access to the database must be documented. A role-based access is advised when any user to a certain role of the database obtain certain access. This also applies for the supporting operating system. Any FOT must define the roles and permissions of the database. These roles can be:

- *Database administrator (DBA):* Unrestricted access to the database.
- *FOT database owner:* Unrestricted access to FOT database data and permitted distribute role access to users.
- *Uploader:* Allowed to insert and update data in the FOT database.
- *Analyst:* Allowed to read data from the FOT database and to manipulate data in private user space.
- *Publisher:* Permitted to insert/update/delete data in shared user space.

- *Web application:* Permitted to read data from specific user space containing aggregated data.

8.2.1.7 Personal integrity and sensitive data

Driver data must be stored according to the access restrictions defined by the steering committee. In a collaborative study, some data may be classified as sensitive by one partner or even by a supplier of measurement equipment.

8.2.1.8 Private vs. public data

Private data should be kept in a private "user space" (database or schema), in order not to risk inadvertent confusion with original project data.

8.2.1.9 Backup

An FOT database backup strategy should be based on "acceptable downtime". Off site backups are mandatory for managing a disaster scenario. The majority of the data is never edited (video and raw data in the database) and data mirroring should be sufficient. For data created by private, organisation, or public user spaces, a daily backup strategy should be applied.

8.2.1.10 Video data (file server)

Please refer to section 5.2.9 in D2.2 for a list of potential standard backup solutions.

8.2.1.11 Database data

The backup policy must be based on the time it takes to recover data and the acceptable loss of data. Even though some studies may use the original logger data as backup, any private or published data created afterwards must have valid continuous backups.

8.2.1.12 Database acceptance

Before an FOT is launched the FOT database architecture should be reviewed by a system evaluator to ensure that all requirements are fulfilled and to verify policy documents.

8.3 Off-line quality management procedures

8.3.1 Quality assurance of objective data

8.3.1.1 Quality assurance before data is uploaded to database

Before uploading objective data from a vehicle, a well-defined algorithm should be applied to all the data in order to verify data consistency and validity.

8.3.1.2 Quality assurance of video data

To catch problems with camera failure or other video related problems, a video checking strategy should be implemented. A tool for viewing one or several images per trip can be useful. Moreover, a function to verify at least the size of video files is necessary — the size is somewhat proportional to recording duration.

8.3.1.3 Driver ID verification/input

Again, it may be necessary to have a process that allows the analysts to view, for example, one image per trip and match this with the IDs of the drivers allowed to drive a specific vehicle. If a driver is unknown, then the data for a particular trip may have to be neglected. A software tool for doing this manual identification of drivers is preferable. Be advised that some eye trackers (if available) provide DriverID functionality.

8.3.2 Quality assurance of subjective data

In order to address the validity of the data, the formulation of the questions (and possible answers) is a key issue, especially when designing a questionnaire to be distributed to respondents. Questions must evidently be formulated in a clear and unambiguous way. In addition, questions must also, e.g.: be specific, not too complicated, be formulated in simple terms that can be understood by the interviewee. Hypothetical questions are the most difficult questions and should be avoided.

Regardless of data source, missing data is a threat to the quality (see Chapter 9). In the case of a missing questionnaire, efforts must be made to ensure that data collection is as complete as possible and reminders must be administered. Furthermore, the number of questions should be thought through, in order to limit the number of questions. In addition, the number of open questions should be as few as possible in order to reduce the effort of the respondents. The interviewer plays an important role in collecting data in an interview situation. Interviewer bias, that is the influence of the interviewer on the respondents' response, can be avoided by administering a questionnaire. However the interviewer may also increase the quality of the data collected by, for instance, answering to questions and using probing questions.

8.3.3 Measures naming guidelines

It is recommended that the FOT project decides on and adheres to a set of naming conventions for measurements. The strategy used should be well documented and thoroughly enforced. Motivations for a clear naming convention include: 1) project-wide consistency, 2) clarity for direct understanding of used measures in analysis, 3) differentiation of non-comparable measurements, and 3) avoidance of confusion.

When specific measurements are named, references to the following measure attributes are recommended: *indicative name*, *associated source*, *sample rate*, and any other FOT *specific descriptor*. The compounds should be joined consistently to create a single word. Possible strategies are: "camel case" (SomeSignal), underlines (some_signal), or hyphens (some-signal). Depending on context and FOT specific requirements all or only a subset of the compounds can be used.

Examples: [GroundSpeed_GPS_1Hz], [GroundSpeedGps1Hz]

The aim is to clearly understand what a measurement "is", where it comes from, and how it relates to other measurements. To avoid the risk of making faulty comparisons, measurements that are *non-comparable* should be named *differently*.

8.3.4 Automatic pre-processing and performance indicator calculations during data upload

It is recommended to define procedures and implementation schemes on *how* to add calculation of pre-processing and performance indicators[FW] in the upload process (see Chapter 9). These calculations should preferably be read-only for the users. The actual algorithms for the pre-processing and performance indicator calculations in this step have to be well defined and tested (on for example pilot test data), or based on previous experience. Since the estimation of some specific performance indicator may set specific requirements on the raw data (see Chapter 5), these constrains have to be taken into account when implementing the automatic pre-processing.

8.4 Data analysis tools

The focus of this section is to describe analysis tools, not to describe analysis procedures or methods.

8.4.1 Data classification/transcription

The exact coding scheme/syntax for events (time segments) will vary widely across FOTs. However, the following features have been identified as important software functionalities:

- organising or categorising subjects into groups and subgroups;
- defining any set and structure of codes, and associating software buttons and keyboard keys to each category;
- editing or updating the coding scheme;
- defining events as either a *state* event[FW](e.g. glance left, glance right) or a *point* event(e.g. stop light);
- defining if state events are mutually exclusive or start/stop and set a default state;
- defining if codes are nominal (e.g. road types) or rating scales (e.g. observer ratings of drowsiness);
- defining if codes are compulsory or optional, logging freely written comments created by human analyst (no coding scheme); and
- support for inter- and intra-rater reliability analyses.

8.4.2 Time history and event analysis with video

This section describes the basic functionality of tools for viewing numerical time history data and the associated environment sensing data which includes video data, map data (e.g. GPS), and traffic state data (e.g. radar).

8.4.2.1 Recommended functionalities for visualisation and interaction with data

- replay single-participant data (numerical time-history data, video data, map data, and traffic state data) simultaneously. Multiple windows for different plots and illustrations provide maximal flexibility to arrange and resize is often spread out on multiple computer screens. Recommended visualisation functionality:
    - Video recordings synchronised with other raw data plots;
    - Continuous variables and performance indicators which can be plotted (and zoomed) on graphs;
    - General information (FOT reference, subject ID, event lists, etc.);
- aggregate and visualise multiple participants' data at once to compare flows of events.

8.4.2.2 Recommended functionality to support data analysis

- database query functionality (e.g. SQL)
- signal processing of numerical data (see also Chapter 9)
- fully customizable mathematical computation, analysis, and algorithm development functionality, automatic or semi-automatic calculation of performance indicators, and application of trigger algorithms to find events of interest (e.g. lane changes, near crashes, jerks)
- image processing of video data (e.g. machine vision algorithms to detect traffic signal status)
- grouped analysis of data (e.g. scripts)
- export results function to tabular format or statistical packages.

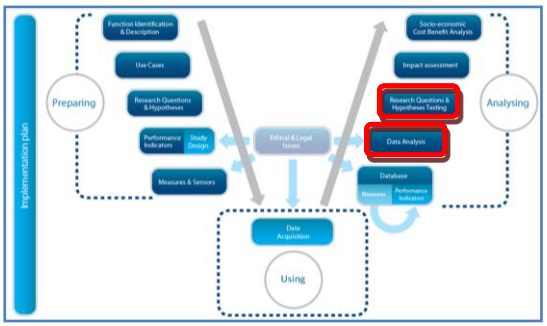
A general recommendation for an analysis package is to use SQL software for database queries, mathematical analysis software for computation (such as Matlab), and common statistical software packages (such as SPSS). If huge datasets have to be analysed or more specific requirements exist, then more specialised or proprietary solutions may be necessary. SQL and some software tools may require a fairly high level of knowledge to use, so it may be advantageous to develop proprietary easy-to-use graphical user interfaces.

8.5 Database usage

Further usage of the database in other FOTs can be considered. In this case, data availability should be addressed. Data acquisition should also take into account this eventual further usage of the database so additional data which might be useful for other FOTs can be selected to be stored, reducing time and cost constraints for upcoming FOTs.

9. Data Analysis and Modelling

9.1 Introduction



The strategy and the steps of data analysis need to be planned in order to provide an overall assessment of the impact of a system[FW] from the experimental data. Despite the appearance of Data Analysis as one of the latter stages in the FESTA V, experts have stressed through FOT-Net 2 that planning for the later analysis work should be done from the start and analysts need to be involved in the research questions to ensure the integrity of the study. Data analysis is not an automatic task limited to some calculations algorithms. It is the place where hypotheses[FW], data and models are confronted. There are three main difficulties:

- the huge and complex amount of data coming from different sensors including questionnaires and video, that needs to be processed;
- the potential bias about the impact of the system(s) on behaviour which may arise coming from sampling issues including location of the study, the selection of a relatively small sample of drivers, etc.;
- the resort to auxiliary models such as simulation models to extrapolate from the behavioural effects estimated and tested within the sample to effects at the level of the whole transport system.

To be confident of the robustness of the outputs of the data analysis, one has to follow some strategic rules in the process of data analysis and apply to the whole chain and to its five links (Figure 9-1) the required techniques such as applying appropriate statistical tests or using data mining to uncover hidden patterns in the data[14].

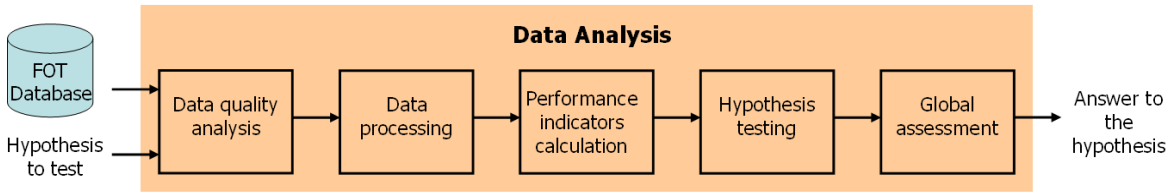


Figure 9-1: Block diagram for the data analysis

Some specific actions are required to tackle the difficulties mentioned above and to ensure the quality and robustness of the data analysis:

[14] For more detailed information the reader should refer to FESTA Deliverable D2.4.

1. A pilot study is a prerequisite to check the feasibility of the chain of data collection and evaluation and to achieve a pre-evaluation of the usefulness of the system.

2. The data flow has to be monitored in detail but also overall. One of the strategic rules to follow is to ensure local and global consistency in the data processing and data handling and analysis.

3. The sources of variability and bias in the performance indicators[FW] have to be identified, where feasible, in order to control for them in the data analysis.

4. There is a crucial need for an integrative assessment process which should ideally combine within a meta-model information gathered on the usability, usefulness and acceptability of the system with the observed impacts of the system on behaviour. The estimated effects obtained from the sample of drivers and data have to be extrapolated using auxiliary models to scale them up.

5. Appropriate techniques have to be applied for each link of the chain — data quality; data processing, data mining and video analysis; performance indicator calculation; hypothesis testing; and global assessment. The techniques come from two set of statistical and informatics tools belonging to two main kinds of data analysis: exploratory (data mining) and confirmatory or inferential (statistical testing).

9.2 Large Data-set handling

An FOT often collects so much data that there are not enough resources and time to analyse all data in the timeframe of the FOT project. There are different choices when it comes to selection of data for analysis.

An option is to take the "space mission" approach in which as much data as possible are collected, because the FOT provides a unique opportunity (and funding) to collect data which may be hard to collect later on. This approach gives a rich dataset, that enhance the probability that the data will be re-used in future projects. However, before starting data collection, it is recommended to develop a plan on how to store the data and how to make it available for later analysis or analysis by others. This plan should specify detailed data dictionaries, open software formats, rules for data access and other relevant information as meta-data.

Analysis later on and by others (in other words, re-using data from other projects) is a good idea, thus reducing the need for expensive and time-consuming data collection phase. The researchers should though be aware that data may become out of date, because traffic, vehicles, driver support and information systems change. Therefore data, which is collected today, might not be of much relevance in ten years time, because of the changed environment and driver behaviour. However, although the context may change, the fundamentals of driving behaviour do not. Therefore whether it is possible to re-use data fruitfully, depends on what is wanted to be known about driving with a support of information system. New projects should be aware that sponsors and stakeholders may want to have fresh data, but the interest for global re-use of data is increasing.The opposite approach is to collect only a minimum set of relevant data or to trigger data collection for

the specific events of interest. Limiting data to specific events may have the consequence that it is not possible to look at generalised behavioural side-effects. Selection of data should be driven in the first place by the research questions[FW] that needs to be answered. With limited resources it may be useful to find a compromise between an explorative study with naturalistic driving and a more strict experimental study in which the expected behaviour of drivers and systems are evoked in a more condensed manner, requiring less time and providing more focussed data. Usage of this selected data for other purposes and projects might not be feasible as the selected data has been collected for certain research questions. Even for the later analysis the specification of the relevant data can be changed (e.g. threshold for an event[FW]) because of new findings within the analysis. An adaptation of these selected data will not be possible, because of missing data.

To make analysis more efficient, it is recommended to take a layered approach to data analysis, making sure that first those data are selected that are needed to provide information on the research questions before going into a detailed analysis. Moreover it needs to be checked, whether the selected data are appropriate to perform the analysis before starting the actual data analysis.

It should be kept in mind that, very often, less is more, since in the end the complete data set cannot be analysed due to

- delays,
- missing data,
- bad data quality,
- budget restrictions,
- limited time,
- restricted access to gathered data in database

The lack of resources to analyse all data is usually the lack of human resources, and not a problem of computational resources. Thus methods for automation of the analysis are needed in order to increase especially the processing of data (e.g. recognition of events[FW]). The analysis of video data is generally a time consuming task, which should be considered from the beginning with respect to planning. Data mining methods are important to tackle this problem. An additional problem with resources is that data analysis comes late in a project. If delays occur in the data collection phase, which is often the case, the phase of data analysis may have to be shortened and resources will be diminished. It is therefore important to plan the data-analysis from the beginning of the project.

The processed data for analysis is generally stored in databases. The performance of the databases decreases with the amount of stored data. Thus intelligent approaches on data storage need to be applied, in order to avoid unnecessary processing time. Data sets for the analysis may be defined in advance as part of the data acquisition scheme and then processed before storage into the databases.

9.3 Consistency of the chain of data treatment

There will be a lot of computations and data flows starting from the measurements collected into the database through estimation of performance indicators to the testing of hypotheses and on to the global assessment. This process, in the form of a chain of operations, has to be monitored in detail but also overall. There are five operations linked together in terms of data treatment. In addition, three kinds of models are needed as support to carry out the three top operations: probability models for justifying the calculations of the performance indicators, integration models to interpret in a systemic way the results of the test, and auxiliary models to assess the effects on a larger scale (scaling up). Moving from the data to an overall assessment is not only a bottom-up process; it also has to include some feedbacks (Figure 9-2). There are two movements along this chain: a data flow going up and a control feedback loop from the top which concerns the consistency of the evaluation process and which mainly depends on control of uncertainty.

In moving up the chain, the consistency of each operation can be checked locally according to the specifications which are governed by the nature of the performance indicators which correspond to a set of hypothesis related to the use cases of the system. For each PI, there are some rules which ensure the validity of the calculation procedures. For example, it is important to sample data which can change rapidly at a high data rate. The sampling rate must fit the variability of the variable. From a database design point of view, however, it may be easier to collect relatively static data at a high frequency.

As a complement to local consistency, a global criterion is to have sufficient sample size to get enough power to carry out the test of a hypothesis or to make an overall assessment with enough precision. This is a feedback loop coming from the top to control the uncertainty of the estimations. The precision required for measurements depends on the uncertainty of the auxiliary models, of the regression models and of the probability models.

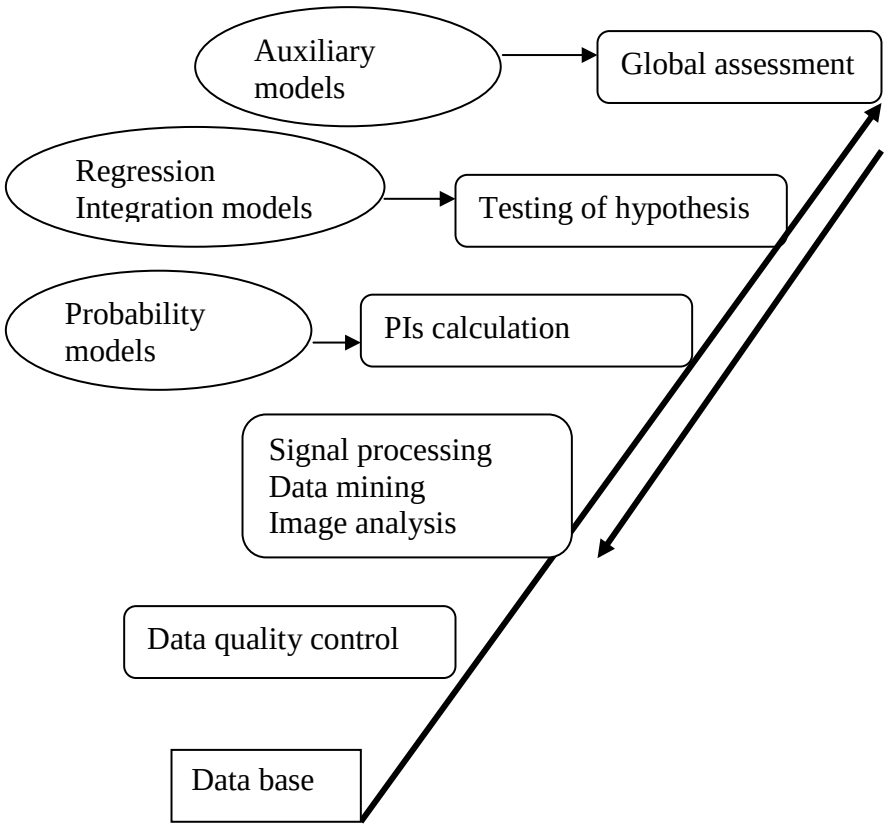


Figure 9-2: Deployment of the chain with feedbacks and additional models

9.4 Precision in sampling

The aim is to measure the effect of an intervention or treatment — which in the case of an FOT is the use of a system or systems — on a sample of subjects and in various driving situations while controlling for external conditions. From the sample, we have to infer the effect on the population by aggregating the values obtained through the sensors without and with the system to get an estimate on the effect on the chosen PI. How to insure that this inference is valid, in other words that the estimation is very near the true effect in the population? The precision of the estimate depends on the bias and variance which could be combined to get a measure of the sampling error (Wannacott and Wannacott, 1990). To control the bias and variance, one has to rely on a well defined sampling plan using appropriate randomisation at the different levels of sampling: driver, driving situations and measurement.

Consideration should be given to identifying the possible sources of (unintended) bias and variance in the sample and either attempt to minimise or account for these in the data analysis. This is one of the most fundamental principles of statistical methods.

1. Driver variation. The simple fact of the matter is that drivers vary. The range of behaviours that drivers exhibit (in terms of speed selection, headway preference, overtaking behaviour) is immense, but fortunately the variation obeys some probability laws and models. Strict randomisation procedures ensure that only the

outcome that is being varied (or the outcome whose variation we are observing) is working systematically. However, strict randomisation is not usually possible or desirable[15] in an FOT. particularly when the sample sizes are relatively small. The theoretical best method is to stratify the population of drivers according to some variables or factors related to the outcome and to sample proportionally to the size of the sub-population and to the a priori variance of the outcome (e.g. speed choice). For practical reasons, a different sampling or selection procedure may be followed. In either case, it is important to be able to compare the sample to the overall driver population in order to identify what are the main discrepancies and to assess possible sources of bias.

2. Driving situation variation. There will be variation within and between the journeys and the driving situations within these journeys. For example a particular journey may be affected by congestion part-way through, or weather conditions may change from day to day. This type of variation cannot be controlled and is considered to be random. The observation period should be sufficiently long to allow for these random effects. One example here is that seasonal effects should be considered.

3. Measurement variation. Once in a driving situation, by means of the sensors, we get a series of measurements at a certain frequency. Their size is not fixed but varies. Each set of measurements within a driving situation constitutes a sample of units taken from a cluster, according to sampling theory. Usually, there is a correlation between the measured outcomes. The information coming from this sample of measurements is not as rich as expected from an independent sample. One such cluster is at the driver level — the data collected from one driver is not independent.

How to quantify the variance of the estimate of an outcome from the experimentation taking into account theses three sources of variations? The total variance of the average of the indicator on the sample breaks down into an inter-individual, intra-individual and infra-situational variance. If the inter-individual variance is strong, an increase in number of situations observed and in the measurement points per situation will not bring any precision gains (Särndahl et al., 1992). However, it may help to ensure a reduction in bias from, for example, seasonality.

9.5 Requirements for integration and scaling up

Having treated and aggregated the data by means of statistical models, there are two kinds of problems to solve related to first the synthesis of the outputs and second to the scaling up of the results from the sample to a larger population. Integration of the outputs of the different analysis and hypothesis testing requires a kind of meta-model and the

[15] It may not be desirable, for example, to waste sample size by recruiting drivers who only drive small amounts each week. Many FOTs have for good reasons used a quota sampling procedure, in which equal numbers of (say) males and females are recruited. This can create bias when scaling up the observed data to estimates of effects at a national or European scale.

competences of a multidisciplinary evaluation team (Saad, 2006). Scaling-up relies upon the potential to extrapolate from the performance indicators to estimates of impact at an aggregate level.

It is often necessary to employ quantitative models from previous studies to estimate the effect of indicator in question. It is, however, important to note that individual models have usually been developed for particular purposes, from particular data and with specific assumptions. However, in the absence of appropriate models available for the purpose of study, it is usually necessary to apply the "least bad" model available with appropriate weighting or adjustment.

It is also important to consider the constraints, assumptions, and implications behind the design of the study in mind when interpreting the analysis results. Behavioural adaptation may lead to side effects (i.e. indirect effects) and also result in a prolonged learning process. However, the study period may, for practical reasons, not be sufficiently long to fully explore this.

Extrapolating from the sample to the population depends on the external validity of the experiment. The power of generalisation to the population of the estimates of impact is related to their precision which is composed of two parts — bias and variance. We can use three approches:

1. If the required performance indicator is available in the sample (e.g. if journey time is an impact of choice for efficiency and journey time has been collected), the impact at the population level can be calculated directly, although sometimes a correction factor or other form of extraplotaion adjustment may have to be introduced (Cochran, 1977).

2. If neither a performance indicator nor a proxy indicator is available, then it is necessary to adopt an indirect approach through models which provide an estimate of the output from the behavioural performance indicator estimated from the sample. Speed changes can be translated into changes in crash risk by applying statisticaly derived models from the literature which have investigated the relationship between mean speed, speed variance or individual speed and crash risk. Emissions models can be used to calculate the instantaneous emission of a car as a function of its recorded speed and gear selected.

3. Finally a macroscopic or microscopic traffic simulation model can be applied to translate the effects observed in the sample to a network or traffic populations effect. The outputs from such a simulation can for example, be used to calculate journey time effects or fuel consumptions effects at the network level.

For scaling up, the sample that is chosen for the FOT is very important. For example, if the FOT was carried out with mainly male participants between the age of 25 and 45, the results can in principle not be extrapolated directly to the whole population of drivers. Getting a representative sample of the whole population is impossible. However, it is acceptable to have an imperfect sample, as long as the limitations of the sample are known and they are

described in the end results. It is still desirable to try to have a sample as representative as possible (gender and age are important).

More detailed information about scaling up can be found in the next section.

In the Amitran project work has been done on scaling up (methodology and data collection). More information about this can be found in (Mans et al., 2013).

9.5.1 Scaling up methods

There are two methods of scaling up that can be used . The first method is a direct method, using statistical data information. The second method is performed through modelling using a macroscopic (multimodal) traffic model on EU 27 level. The choice of scaling up method is based, among others, on the availability of models and the type of effects expected.

The methods described in the following explain how scaling up can be applied theoretically. In practice, scaling up is a big challenge. It is important to consider the goal one wants to achieve.

9.5.1.1 Scaling up using statistics

The scaling up method using statistics, initiates from the impacts on CO_2 emissions at a local level as distinguished for different situations (such as traffic state, vehicle type, etc.), coming from the FOT. In case it is not possible to directly use the local effects of the system to scale up with the use of appropriate statistical datasets, then the use of models (e.g. microscopic traffic models) is necessary to transpose this impact to a more appropriate format for scaling up. The definition of situations depends on:

- the system characteristics
- the situational variables that are expected to have the largest impact (e.g. a night vision system will only be active during driving in the dark)
- the possibility of measuring the different situations and the model capabilities.

Data for the same situations is needed on the large scale level that is targeted. Then, the impact on a local scale is scaled up using statistical data (for example on kilometres driven for the different road types) under the specific situations.

Scaling up using statistics is applicable when interaction and second-order effects (i.e. latent demand induced by the improvement of the service level, caused by a system) can be expected to be insignificant, or when there is a clear effect at certain traffic situations for which data on higher level are available, or even at the mere event that no appropriate macroscopic model is available to perform the model-based methodology. A drawback of this method is that data sets need to be available for the countries one wants to scale up to. At present there is very limited measurement data for some countries in Europe plus, a (software) tool for this approach does not exist yet. The Amitran project is collecting statistics for scaling up from various European countries in a knowledge base. This knowledge base will be made public around Summer/Autumn of 2014.

9.5.1.2 Scaling up using a macroscopic (multimodal) traffic model

The network of a macroscopic (multimodal) traffic model determines the level on which the results are calculated. Ideally the model is available on country or EU level. Scaling up using such a model can be done in two different ways:

- The calculation of the impact is done with a model other than the macroscopic traffic model. The local effects of the system are in this case determined (e.g. via a microscopic simulation tool). These effects can be used as input for the macroscopic model on country/EU level. One run is performed for deriving the direct effect to a larger scale.
- The calculation of the impact is performed directly with a macroscopic traffic model. In this case, should the model be at the required level (country/EU), the direct effect of the system is calculated. This can be done performing a run of the macroscopic model. A limitation to this approach is that microscopic effects of ITS cannot be taken into account, e.g. changes in driver behaviour. Therefore it can only be used to determine the effects of ITS that mainly affect macroscopic mechanisms in the network, such as mode or route change.

Optionally (for both cases), the economic effect can be calculated with an appropriate model. Then a second run is performed with the macroscopic model to account for the second order effect.

Scaling up using a macroscopic model is a good method to apply when second-order effects are expected and/or when the effects of the ITS system can be used directly as an input parameter for the macroscopic model. Also, this method can be used only if such a large-scale model is available. Being a more elaborate method than scaling up using statistics, it allows taking into account specific circumstantial differences especially if there are interaction effects. A downside of scaling up with a macroscopic traffic model is that urban roads are usually not part of the network on such a large scale, and also, that it requires more effort than scaling up using statistics.

9.6 Appropriate techniques at the five links of data analysis

The five links follow the right branch of the development process of an FOT from data quality control to global assessment. Different techniques of data analysis and modelling which could be used at each step are presented here.

9.6.1 Step 1: data quality analysis

Data quality analysis is aimed at making sure that data is consistent and appropriate for addressing the hypothesis of interest (FESTA D3, section 4.5). Data quality analysis starts from the FOT database and determines whether the specific analysis that the experimenter intends to perform on the data to address a specific hypothesis is feasible. Data quality analysis can be performed by following the four sub-steps reported below (and shown in Figure 9-3). A report detailing the quality of the data to be used to test the hypothesis of interest should perhaps be created.

The sub-steps for data quality analysis are:

1. **Assessing and quantifying missing data** (e.g. percentage of data actually collected compared to the potential total amount of data which it was possible to collect).

2. **Ensuring that data values are reasonable and units of measure are correct** (e.g. a mean speed value of 6 may be unreasonable unless speed was actually recorded in m/s instead of km/h).

3. **Checking that the data dynamic over time is appropriate for each kind of measure** (e.g. if the minimum speed and the maximum speed of a journey are the same, then the data may not have been correctly sampled).

4. **Guaranteeing that measures features satisfy the requirements for the specific data analysis** (e.g. in order to calculate a reliable value of standard deviation of lane offset, the lane offset measure should be at least 10s long; additionally, this time length may depend on the sampling rate — see AIDE D2.2.5, section 3.2.4).

Please, notice that the first three sub-steps refer to general quality checks; thus, if any of these fails, data analysis cannot proceed. If a failure is encountered, it should then be reported to those responsible for the database responsible so that the possible technical error behind can be tracked down and solved. However, the last sub-step is different, and is related to the specific analysis or to a specific performance indicator to be used in the subsequent data analysis. As a consequence, if step 4 fails, it may not be due to a technical issue that needs to be solved, but to intrinsic limitations in the collected data.

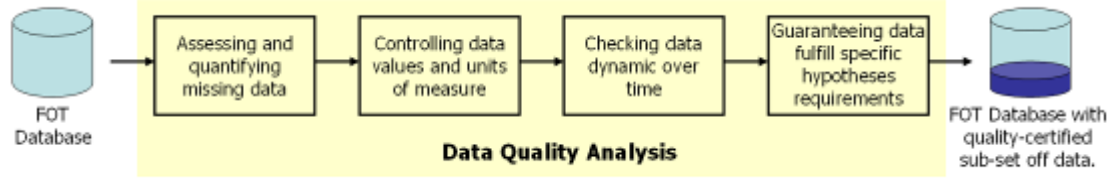

FOT Database
Assessing and quantifying missing data
Controlling data values and units of measure
Checking data dynamic over time
Guaranteeing data fulfill specific hypotheses requirements
Data Quality Analysis
FOT Database with quality-certified sub-set off data.

Figure 9-3: Block diagram of data quality analysis

Data quality analysis is handled differently with regard to data from in-vehicle sensors (generally CAN data and video data) and subjective data (generally from questionnaires). Subjective data, once collected, is hard to verify unless the problem stems from transcription errors.

9.6.2 Step 2: data processing

Once data quality has been established, the next step in data analysis is data processing. Data processing aims to "prepare" the data for addressing specific hypotheses which will be tested in the following steps of data analysis. Data processing includes the following sub-steps: filtering, deriving new signals from the raw data, event annotation, and reorganisation of the data according to different time scale (Figure 9-4). Not all the above-mentioned sub-steps of signal processing are necessarily needed for all analyses. However, at least some of them are normally crucial.

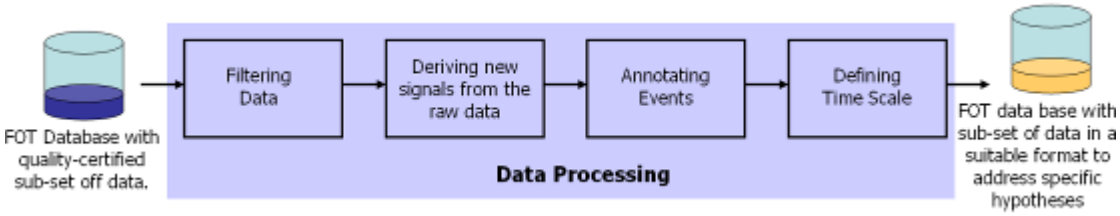

FOT Database with quality-certified sub-set off data.
Filtering Data
Deriving new signals from the raw data
Annotating Events
Defining Time Scale
Data Processing
FOT data base with sub-set of data in a suitable format to address specific hypotheses

Figure 9-4: Block diagram for the procedure of data processing

Data filtering can involve a simple frequency filter, e.g. a low-pass filter to eliminate noise, but also any kind of algorithm aimed at selecting specific parts of the signals. Very often a new signal more suitable for the hypothesis to be tested has to be elaborated by combining one or more signals. Marking specific time indexes in the data, so that event of interest has been recognized, is fundamental to individuate the part of data which should be analysed. Ideally, an algorithm should be used to go through all FOT data and mark the event of interest. However, especially when the data to be annotated is from a video and requires the understanding of the traffic situation, writing a robust algorithm can be very challenging even with advanced image analysis techniques and manual annotation from an operator may be preferable. Re-organising data into the most suitable time scale for the specific hypothesis to be addressed has to be considered in the following steps of the data analysis.

9.6.3 Step 3: Performance Indicators calculation

There are five kinds of data which provide the performance indicators: Direct Measures, Indirect Measures, events, Self-Reported Measures and Situational Variables. The scale of the dataset and the uncontrolled variation in driving situations that occurs from driving freely with vehicles become a seriously limiting factor unless efficient calculation methodology is implemented. The choice of which performance indicators and hypotheses to calculate is clearly dependent on the amount of effort required. Efficient calculation methods need to anticipate that (a) performance indicators will be calculated on imperfect data – there is a strong need to create special solutions for “exceptions to perfect data”, and (b) performance indicators calculation requires situation or context identification — a “denominator” or exposure measures to make a measure comparable is required to determine how often a certain event occurs per something (e.g. km, road type, manoeuvre). The fact that test exposure is largely uncontrolled (not tightly controlled as in experiments) means that analysis is largely conducted by first identifying the important contextual influences, and then performing the analyses to create a “controlled” subset of data to compare with.

The ability to find and classify crash-relevant events (crashes, near-crashes, incidents) is a unique possibility enabled by FOTs to study direct safety measures. As reported in 5.3.3, this possibility should be exploited by using a process of identification of critical events from review of kinematic trigger conditions (e.g. lateral acceleration >0.20 g). The definition of these trigger values and the associated processes to filter out irrelevant events are of particular importance for enabling efficient analyses.

Care should be taken to use appropriate statistical methods to analyse the performance indicators. The methods used must consider the type of data and the probability distribution governing the process. Categorical or ordinal data, such as that from questionnaires, needs to be analysed appropriately. Data on the degree of acceptance of a system (e.g. positive, neutral, negative) can be applied in multivariate analysis to link it to behavioural indicators so as to create new performance indicators.

9.6.4 Step 4: hypothesis testing

Hypothesis testing in an FOT generally takes the form of a null hypothesis: no effect of the system on a performance indicator such as 85th percentile speed, against an alternative such as a decrease of x % of the performance indicator. To carry out the test, one relies on two samples of data with/without the system from which the performance indicator is estimated with its variance. Comparing the performance indicators between the two samples with/without intervention is done using standard techniques such as a t-test on normally distributed data. Here the assumption is that there is an immediate and constant difference between the use and non-use of the system, i.e. there is no learning function, no drifting process and no erosion of the effect.

However, the assumption of a constant effect is often inappropriate. To get a complete view of the sources of variability and to handle the problem of serially correlated data, multi-level models are recommended (Goldstein, 2003).

With such models, drivers or situations with missing data have generally to be included. Elimination of drivers or situations because of missing data in order to keep complete data set may cause bias in the estimation of the impact.

It is assumed that data will have been cleaned up in the data quality control phase. Nevertheless, to be sure that the estimation will be influenced minimally by outliers, one can use either robust estimates such as trimmed mean and variance or non-parametric tests such as a Wilcoxon rank test or a robust Minimum Mean regression (Gibbons, 2003; Wasserman, 2007; Lecoutre and Tassi, 1987). Such tests provide protection against violation of the assumption of a normal distribution of the performance indicator.

9.6.5 Additional Step 4: data mining

Data mining techniques allow the uncovering of patterns in the data that may not be revealed with the more traditional hypothesis testing approach. Such techniques can therefore be extremely useful as a means of exploratory data analysis and for revealing relationships that have not been anticipated. The data collected in an FOT is a huge resource for subsequent analysis, which may well continue long after the formal conclusion of the FOT. One relatively simple technique for pattern recognition is to categorise a dataset into groups. Cluster analysis tries to identify homogeneous groups of observations in a set of data according to a set of variables (e.g. demographic variables or performance indicators), where homogeneity refers to the minimisation of within-group variance but the maximisation of between-group variance. The most commonly used methods for cluster analysis are k-means, two-step, and hierarchical clusters (Lebart et al., 1997; Everitt, 2000).

9.6.6 Step 5: global assessment

This section deals with the issue of identification of models and methodologies to generalise results from a certain FOT to a global level in terms of traffic safety, environmental effects and traffic flow. One problem when generalising results from an FOT is to know how close the participants in the FOT represent the target population. It is often necessary to control for: usage, market penetration and compliance (the system might be switched off by the driver) and reliability of the system. The process of how to go from the FOT data to safety effects, traffic flow and environmental effects is illustrated in Figure 9-5. In this process two steps need to be taken. One is scaling up the FOT results, for example to higher penetration levels or larger regions. The other is to translate the results from the level of performance indicators (for example, time headway distribution) to the level of effects (for example, effect on the number of fatalities). For each type of effect there are (at least) two different ways to generalise the results: through microsimulation or directly.

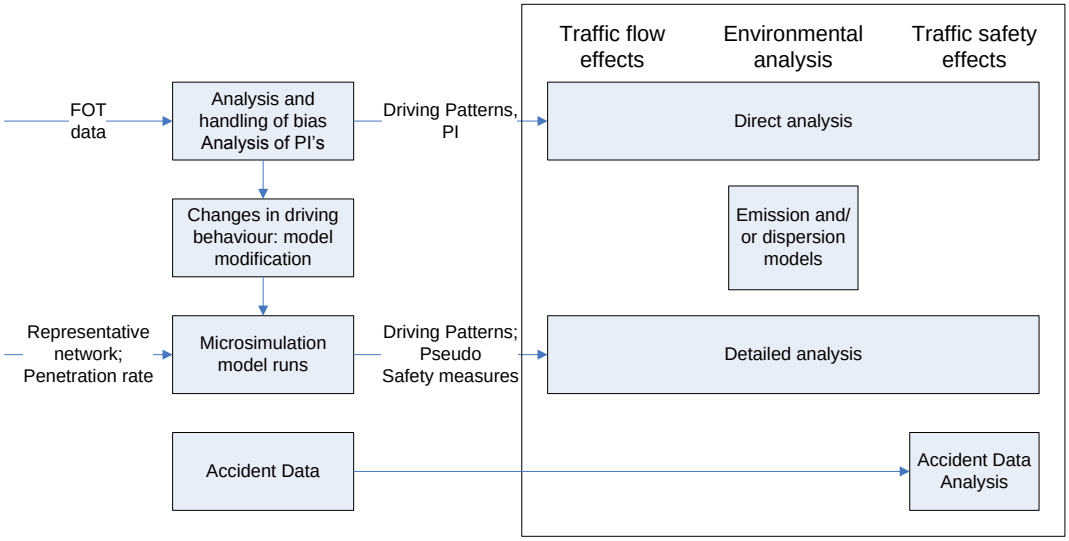


Figure 9-5: Block diagram of translating FOT indicators to large scale effects

The direct route includes both estimation directly from the sample itself and estimation through individual or aggregated models. Some advantages of the direct route are that it is rather cheap and quick. The alternative is to use a traffic microsimulation model which represents the behaviour of individual driver/vehicle units. The advantages of microsimulation are that they can be more reliable and precise and can incorporate indirect effects (such as congestion in the network at peak times).

The simulation of indirect effects allows studying interaction effects between equipped and non-equipped vehicles. This aspect is of major importance when testing a certain function in an FOT. By means of the collected real world data only direct effects caused by the tested function in the equipped vehicle can be analysed. Effects due to interaction with other vehicles cannot be analysed, because data collection is only carried out by means of the test vehicles. Hence traffic simulations are a useful tool to investigate indirect effects in different driving situations. The indirect effects are also of interest when testing connected vehicles. The interactions between connected and non-connected vehicles provide necessary

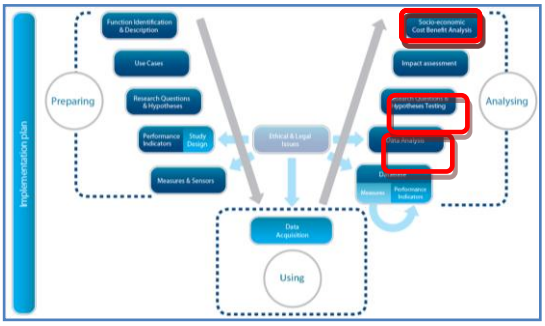
information for the assessment of safety and traffic flow effects. However, microsimualtions for connected vehicles require further information to provide a model for the simulation environment. Information on the communication range of the connected vehicles, information on when vehicles are connected to another vehicle, and the integration of the communication infrastructure might require more efforts as originally foreseen.

Microsimulations also allow studying certain effects or functionalities before the data collection phase starts, in order to get an idea of the function and its effect. Here different settings of the function can be tested. Another possibility to get an idea of the effects in advance is the conduction of driving simulator tests. These tests can be performed in parallel or before the real world tests. In comparison to microscopic simulations, driving simulator tests require higher efforts. Especially the participant recruitment, the conduction of interviews after the tests as well as the analysis of the collected data (subjective and objective) require a large amount of time, which needs to be taken into account in the preparation phase. In some cases the simulations might need input from the real traffic, in order to understand how certain situations are evolving. Especially data on the relevant network and the traffic density, e.g. whether the vehicle is driving in free flow or in a car following situation are needed for traffic simulations. But also subjective data is of importance as well. For instance the questions why certain drivers did not follow a certain function warning or advice might be interesting, when implementing the driver model. Interviews with drivers can be an efficient approach to understand certain driver behaviour patterns. Moreover traffic simulations are an important tool for the interpretation of the results or determined effects. By means of the traffic simulations more details on the relevant network can be observed and provided, in order to understand a certain behaviour or effect.

Since traffic microsimulation models consider individual vehicles in the traffic stream, there is consequently the potential to incorporate FOT results in the driver/vehicle models of the simulation. Impacts on the traffic system level can then be estimated through traffic simulations including varying levels of system penetration into the vehicle population. Here it is of importance to provide a detailed driver model. The driver model becomes more realistic if required details are determined by means of data gathered in real traffic. This requires in addition the provision of a driver model for the baseline behaviour as well as the adapted behaviour in the treatment phase, due to the usage of a tested function.

Microsimulation does not necessarily yield the impact variable that is of interest. Various aggregated and individual models are necessary to convert for instance speed to safety effects (e.g. via the Power Model which considers the relationship between driving speed and the risk of an accident at different levels of severity). In addition, the modelling detail of traffic microsimulation places restrictions on the practical size of the simulated road network. Macroscopic or mesoscopic traffic models combine the possibility to study larger networks with reasonable calibration efforts. These models are commonly based on speed-flow or speed-density relationships. Large area impacts of FOT results can therefore be estimated by applying speed-flow relationships obtained from microsimulation for macro- or mesoscopic traffic modelling.

10. Impact Assessment and Socio-Economic Cost Benefit Analysis



10.1 Introduction.

All FOTs[FW] require an impact assessment analysis. Many times, especially if the FOT is supported by funding from the European Commission, the requirement is that the assessment should be performed at European level, also with details on the socio-economic evaluation.

> Naturalistic Driving Studies tend to focus on crash-explanatory factors, and Field Operational Tests generally focus on evaluation of systems[FW] or functions[FW]. Therefore this chapter is of less relevance for NDS, depending on the hypotheses and research questions and whether impacts are involved. Part of this chapter may be relevant for NDS, for example if an overview of the findings has to be presented, or if findings are used to identify measures or in development of functions and services.

In the past, the impact assessment of FOTs focussed on a narrow set of impacts of interest. Few looked at the stakeholder or supplier perspectives; some measured benefits but not (social) costs; very few started out with an impact table and formally identified what the expected "main effects" of the systems investigated would be; and some did not carry out a socio-economic impact assessment.

The goal of this chapter is to provide concise advice on how to carry out an impact assessment. Many parts of the chapter will also be relevant to FOTs conducted at a national or regional level within Europe or to FOTs conducted outside Europe. A consistent methodology for carrying out this analysis for EC-funded FOTs will maximise the comparability of the results across regions.

Another goal is to address the possible breadth of impacts that can be considered and the available resources for carrying out the assessment, like references to examples of good practice in existing (web) documents. Although impact assessment comes at the end of the FOT chain, it should also play a role at the beginning of a project, indeed some of the choices made in setting up the FOT are linked to the Cost Benefit Analysis (CBA). For example, choices about performance indicators and scaling up are directly linked to what can be analysed in terms of scope and impacts in the CBA. Furthermore the expected impact is a major driver for the decisions to be made in the design of the FOT. People who are responsible for the impact assessment should be involved from the beginning of the project.

It should also be considered that policy makers expect concrete results from an FOT, such as the the gain in term of reduction of accidents, and scaled up costs and benefits (on a European level). This is often promised in project proposals. However, these are results that are difficult to get and there could be various reasons why the FOT does not deliver these results in the end (or at least not with the completeness that policy makers desire).

Managing expectations is therefore important, throughout the project, explaining any apparent failures early, whilst emphasising lessons learned or any secondary benefits found.

Our advice will be useful for a variety of parties: the organisations conducting the FOTs, including also the impact assessment specialist; the client commissioning the FOTs; and the consortia drawing up proposals for the FOTs. This chapter assumes that a "professional" in the area of impact assessment and of socio economic cost benefit analysis will carry out the analysis. This individual we will refer to as the "analyst". This information on impact assessment is not meant as a "tutorial".

Broadly, there are three research angles where the evidence comes from:

- **Impact assessment studies** typically investigate the impacts of a system for a future time horizon. These prospective studies make use of an ex-ante impact assessment, often based on literature review, simulation work and expert estimation. They are often comprehensive in scope but they do not involve, or only to a limited extent, data from real-life conditions.

- **Transport appraisal guidelines** or **scoping studies** in this area are very much focused on the appraisal part of the impacts. They dig deep into the methodology and practise of appraisal. They also involve proposals for standardisation of appraisal. Their detriment is that they are not developed for specific use in the field of safety evaluation.

- **Field Operational Test assessment studies** typically assess the impacts of one or more system functions. FOT evidence can lead to a quantum leap in the impact assessment because FOT produce measured data about effects. Therefore, the assessment can rely on ex-post measurement data.

All different research angles have their specific strengths and weaknesses. In preparing this guidance document, we found it useful to combine the strengths of the different perspectives.

10.2 Considerations

The impact assessment investigates the impacts of a technology on society. Ideally an impact assessment provides the decision maker with relevant information in a concise format. The relevant comparison is between the benefits and costs between a base case, e.g, a scenario without the ICT system("without case") compared to those of the scenario with the ICT system ("with case"). In preparing to carry out an impact assessment, the analyst is faced with making choices about the deployment scenarios, the geographical scope of the assessment and the analyses to be carried out. This section will go deeper into the issues surrounding these choices. The chapter will conclude with guidance on how to make the choices and carry out impact assessement and the socio economic cost benefit analysis.

10.2.1 Deployment scenarios

Predictions of the future, particularly over the medium or long term, cannot be precise. This argues for a scenario-based approach when developing forecasts of how future deployment

of a system might turn out. This approach permits alternative scenarios to be evaluated in the cost-benefit analysis. Very likely one scenario will emerge as more favourable with the highest benefit to cost ratio, although that scenario may not be the most probable. The scenario analysis will also enable obstacles, to the pursuit of that scenario, to be identified. This in turn can identify public policy needs and other stakeholder requirements.

For the socio-economic impact analysis deployment scenarios are required, whether they are just implicit or explicit. One approach is not to make any specific predictions but simply examine the socio-economic impacts of different levels of penetration. But even for such a simple set of scenarios, there has to be an assumed growth in penetration over a period of years. But it is more appropriate to build up some alternative scenarios with different "futures". There are a number of potential inputs into creating those scenarios:

1. Policies of public authorities at a European, national , regional and local level. Policies and strategies for transport in general, road safety, environment, accessibility, traffic management and general ICT deployment (e.g. future mobile network capacity) are all relevant.

2. Other quasi-regulation such as EuroNCAP and standards

3. Stakeholder plans — the strategies of OEMs, telecoms operators, road operators, large fleets and so on

4. Likely developments in the various relevant markets, including costs, technical issues (such as synergies between different systems) and competitive pressures

5. Public-private partnerships and their influence on the deployment process

6. The attitudes and willingness to pay of the general public to identify potential purchasing and usage decisions.

Deployment can be pushed or even mandated by the public authorities, which can be termed "regulatory deployment". It can be more voluntary, with a push from major stakeholders, e.g. by formal Voluntary Agreement. Or it can be purely market-driven, i.e. totally voluntary, depending entirely on the public's willingness to purchase systems and use functions.

There are a number of tools for scenario development. At the most basic level there are forecasts on the growth of the vehicle market, changes in mobility and changes in the road network. Government strategies provide information about policy goals and targets. Stakeholder questionnaires and analysis provide further details on willingness to promote and invest. For the views of end-users, feedback from FOT participants on acceptance and willingness-to-pay is an important source of information and should be routinely collected in an FOT. More general information on public attitudes can be collected by means of focus groups, household surveys and stated preference studies. The last are useful as a tool to reveal trade-offs between alternative choices.

10.2.2 General issues of the socio-economic impact analysis

As at the hypotheses[FW] formulation stage, consideration needs to be given to the potential bundles of systems to be handled in the impact assessment. Indeed, there can be a large number of permutations of market penetration of different bundle sizes and not all can be covered, and some combinations of functions may be more likely than others. Some expert judgement has to be applied here, make reference back to market surveys, etc. More information about disentangling the effects of combinations of functions can be found in (Faber et al., 2011).

An impact assessment can only be as good as the data on which it is based. Hence those carrying it out should also be involved in the performance indicator[FW] definition stage in order to ensure that the relevant indicators are being collected. These may be the obvious indicators, such as speed, route choice etc., but there may be occasions where the FOT cannot provide the desired data directly and other methods (e.g. surveys, workshops etc.) may have to be applied.

It is difficult to provide a definitive guide for conduction of an impact analysis. The process depends on the research questions[FW] and the functions. For example in terms of scaling up, the area of interest may be a particular city, country or group of countries. For a function that operates e.g. on highways it makes sense to scale up according to those networks only. There is the added problem that scaling up data from one country to another may be inappropriate for a wide variety of reasons (driver types, weather, cultural aspects etc.). More empirical data and information is needed to increase certainty on effects and conclusions.

A major step of the impact assessment is scaling up by using simulation tools. Simulation can determine indirect effects (i.e. reduction of congestion due to less accidents etc.). Simulation tools generally require modelling of driver behaviour. This modelling relies not only on specific indicators which are being collected in the FOT from the equipped vehicles. Moreover the modelling relies also on information about the interactions of those vehicles with non-equipped vehicles and other road users. The modelling can be performed in micro-simulations, which are able to provide input to the impact analysis. These interactions between non equipped and equipped vehicles are rarely addressed in the hypotheses or in measurement processes. This can only really be addressed by undertaking observational studies (video data analysis could be a possibility).

Current micro-simulation methods have their own limitations. Typically the most generally used software packages do not properly represent vehicle dynamics in terms of interaction with the road surface. They therefore are not properly capable of covering lateral dynamics of the vehicle. The networks covered are often geographically small and have often been created for purposes other than the evaluation of new vehicle-related technologies. They may not be very representative of overall national road networks and origin-destination matrices (i.e. traffic flows) are generally lacking for night-time and weekend periods.

Business analysis in a private environment serves the same purpose as socio-economic impact assessment in the public world. It provides crucial information for use in the decision-

making process on further steps towards deployment. Where socio-economic analysis tries to rule out double counts on the cost side, the business analysis focuses more on the feasibility of the concept and the roles and consequence for each of the actors required in the value web. The process of balancing the transfer of value (hardware, software, money, data, permits etc.) needs to be investigated and usually there are a number of options for the business model (private, public-private and public only), each with a specific constellation of the roles of the actors involved and each with a specific risk pattern.

10.2.3 Assessment scope and process implication

At the start of the socio-economic assessment, a view will need to be taken on the scope of the analysis. Ideally the assessment would include all impacts of the system no matter how small that impact is: safety, mobility, efficiency and productivity, environmental, user acceptance and human factors, performance and capability, legal and implementation issues, and costs. However setting an unlimited such a broad scope for a socio-economic assessment will result in excessive data collection and analysis in terms of expense and time. Given that the purpose of the assessment is to firstly ensure that the implementation of the system is economically beneficial and secondly to aid the choice between alternatives, the scope of the assessment often can be narrowed by excluding minor or insignificant impacts as long as the exclusion of these impacts will not bias the appraisal. An *impact table* such as in Batelle Memorial Institute (2003, p. 45) is extremely useful at the start to clarify which impacts have been considered and which — if any — have been ruled out as negligible or impossible to assess.

10.2.4 Geographical scope of assessment

The issue related to geographical scope is the ability to translate the findings of the FOT to a "higher" geographical level. The FOT is usually carried out at one or more locations, on a regional or national scale. However, the number of equipped vehicles and, if relevant, equipped roads, as well as the number of "equipped" kilometres driven, is usually a small percentage of the total vehicle fleet and the kilometres of roads. Therefore, in order to draw conclusions about the impacts and effectiveness of the system tested, a "scaling up" of the results is needed in order to draw conclusions and in order to ensure transferability of the results. Section 9.5 addresses the scaling up issues, which is to the national or European level. The availability of data plays a role in the decision to what level to scale up the results. Section 10.4 goes into more detail to explain how to deal with this issue.

10.3 Analysis of impacts

The analysis of impacts (impact assessment) represents the most sophisticated part of the assessment. It**Errore. L'origine riferimento non è stata trovata.** provides an overview over the most common effects (safety, mobility, environment, costs) which are considered in an FOT assessment. This assessment framework involves the distinction between direct and indirect effects (in safety mechanisms but also with respect to mobility effects, see below). It also implies the distinction between effects on internal and external costs. Mobility effects typically lead to lower internal costs of transport (i.e. time, fuel consumption) and also external costs (e.g. pollution, CO_2). The reduction of external costs is flagged out separately

under environmental benefits because of its importance on the political agenda. The assessment can of course also consider wider economic effects (e.g. growth and employment effects of new technologies). However, given limited time and budget, it is useful to concentrate on the main impacts.

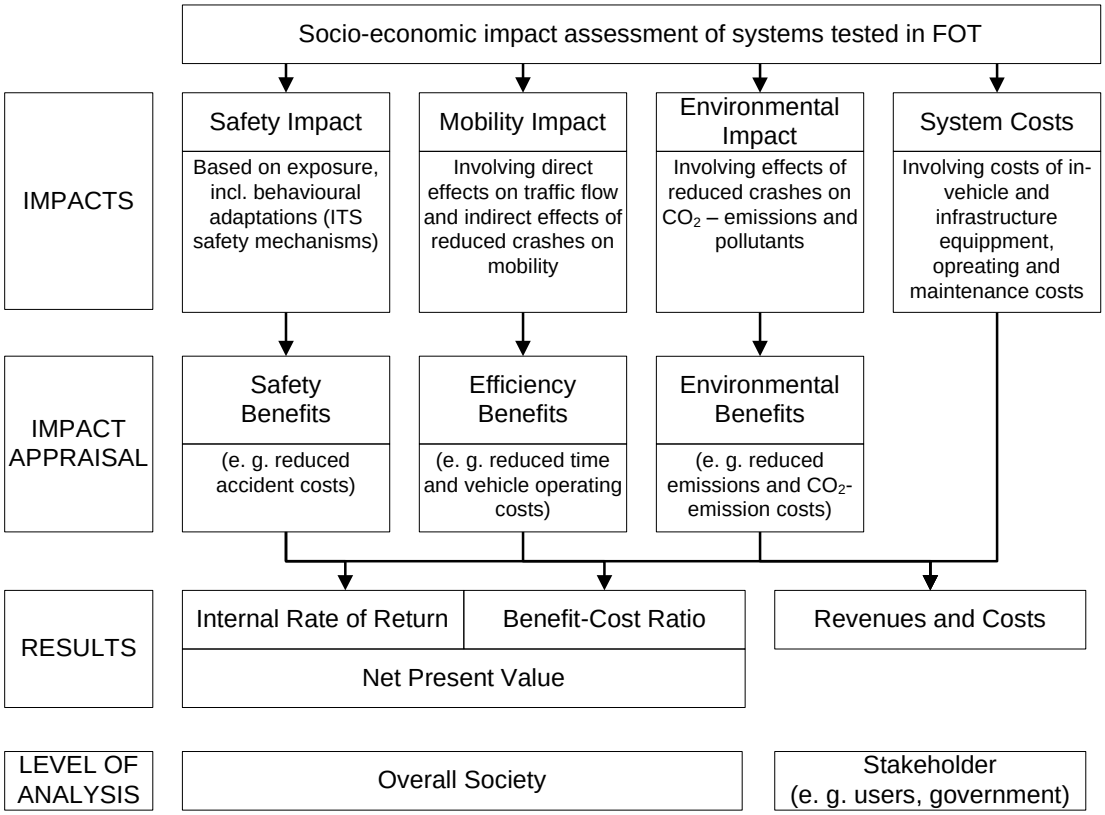

Socio-economic impact assessment of systems tested in FOT
IMPACTS
Safety Impact
Based on exposure, incl. behavioural adaptations (ITS safety mechanisms)
Mobility Impact
Involving direct effects on traffic flow and indirect effects of reduced crashes on mobility
Environmental Impact
Involving effects of reduced crashes on CO_2 – emissions and pollutants
System Costs
Involving costs of in-vehicle and infrastructure equippment, opreating and maintenance costs
IMPACT APPRAISAL
Safety Benefits
(e. g. reduced accident costs)
Efficiency Benefits
(e. g. reduced time and vehicle operating costs)
Environmental Benefits
(e. g. reduced emissions and CO_2-emission costs)
RESULTS
Internal Rate of Return
Benefit-Cost Ratio
Net Present Value
Revenues and Costs
LEVEL OF ANALYSIS
Overall Society
Stakeholder (e. g. users, government)

Figure 10.1: Scope of the impacts within socio-economic impact assessment

10.3.1 Safety benefits

The most direct and easiest way to calculate safety benefits would be to compare the number of accidents (and their consequences) happening during the baseline and treatment phase in an FOT. However, usually not enough accidents happen in an FOT to make this approach feasible. Therefore other methods have to be used. Traffic safety is regarded as a multiplication of three factors, namely exposure, accident risk and injury risk (Nilsson, 2004). The assessment of safety impacts has to consider these three effects which can be combined to predict the overall safety benefit, while taking driving conditions into consideration as well. Strategic decisions are highly relevant for exposure, and driving behaviour (on tactical and operational level) is relevant for accident and injury risk. A change in exposure can be measured in the FOT directly: do people drive more or less with the system, do they drive on other road types, do they choose other routes? A change in mileage has a direct effect on exposure so on the number of accidents. Translating a change in driving behaviour into accident and injury risk is less straightforward. There are gaps in knowledge: the relation between changes in driving behaviour and (number of) accidents is often not known.

Therefore there is not one method that is recommended to use. In this section a number of approaches to calculate the safety benefits of ITS applications are mentioned, with a (brief) explanation, references and information about in what situations they can be used:

- ***Speed (variance) – accident relationships***
  Speed has a close relation to safety. The speed of a vehicle will influence not only the likelihood of a crash occurring, but will also be a critical factor in determining the severity of a crash outcome. This double risk factor is unique for speed. The relationship between speed and safety can be estimated by various models such as the Power Model (Nilsson, 2004; Elvik et al, 2004), that estimates the effects of changes in mean speed on traffic crashes and the severity of those crashes. The Power Model suggests that a 5% increase in mean speed leads to approximately a 10% increase in crashes involving injuries and a 20% increase in those involving fatalities. More examples of models for speed-safety relationships are reviewed in Aarts and van Schagen (2006). In the ISA UK project this is elaborated, different relations (models) for different road types can for example be found in Table 13 in http://webarchive.nationalarchives.gov.uk/20110304132839/http:/cfit.independent.gov.uk/pubs/2008/isa/index.htm .
  The Power Model is valid under the assumption that mean speed is the only factor that has changed. Therefore these models are more suitable for FOTs with systems mainly dealing with speed, and even then they fail to consider changes in the distribution of speed (shape of the speed distribution and changes in speed variance). The model is not suitable for systems that for example influence lateral behaviour.
- ***Event based analysis***
  Crashes are very rare events[FW], thus there is a strong interest and need for the use of crash surrogates or "crash-substitute" events. The basic idea is that less severe events can be used instead of crashes to estimate safety benefits because there is a systematic and well-understood relationship with crashes. Event based analysis (EBA) uses events to estimate safety benefits. As reported in 5.3.3 the basic principle of EBA is to identify short driving segments (typically in the order of 5-10 seconds), during which the risk of crashing is judged to be higher compared to other driving in the data set, and then to analyse these events further. These events are often referred to as Crash Relevant Events (CRE), since their occurrence is thought to be indicative of actual crash risk in one way or another.
  EBA can be used for functions that warn for a certain event (e.g. FCW), not for functions that work continuously.
- ***eIMPACT method***
  The eIMPACT method was used in the eIMPACT project for the safety assessment of in-vehicle safety systems (IVSS). The complete methodology can be found in (Wilmink et al., 2008). In short it works as follows. Effects of IVSS on traffic safety may appear in many, both intended and unintended ways. It is not possible to define in advance the group of accidents affected by the system, although system developers typically have as a starting point a target group of accidents for a system. Therefore, it is highly important that the analysis of IVSS covers all possible effects in

a systematic manner. The approach was based on the system nature of transport. When one element of the system is affected, the consequences may appear in several elements and levels of the system, both immediately and in the long term, due to behavioural modification. Road safety is regarded as a multiplication of three orthogonal factors: (1) exposure, (2) risk of a collision to take place during a trip and (3) risk of a collision to result in injuries or death. In the analyses, the three main factors of traffic safety were covered by nine behavioural mechanisms as first described in (Draskóczy et al., 1998). Five mechanisms are mainly connected to the accident risk, three mechanisms deal with exposure, and there is one mechanism that deals with changes in accident consequences.

Every mechanism may result in either positive or negative impacts on road safety. In summary, the analyses aims to cover not only the direct intended effects of systems but also the indirect and unintended effects, including behavioural adaptation in long term use. In addition, it was taken into consideration that the effects will vary according to road conditions and circumstances. This should ensure that all effects on safety are covered by the analyses.

The starting point for the safety impact assessment were the system specifications, including detailed safety function definitions. Figure 10.2 presents an overview of the phases in the analysis. An important part of the analysis is the use of accident data. For further details the reader is referred to (Wilmink et al, 2008).

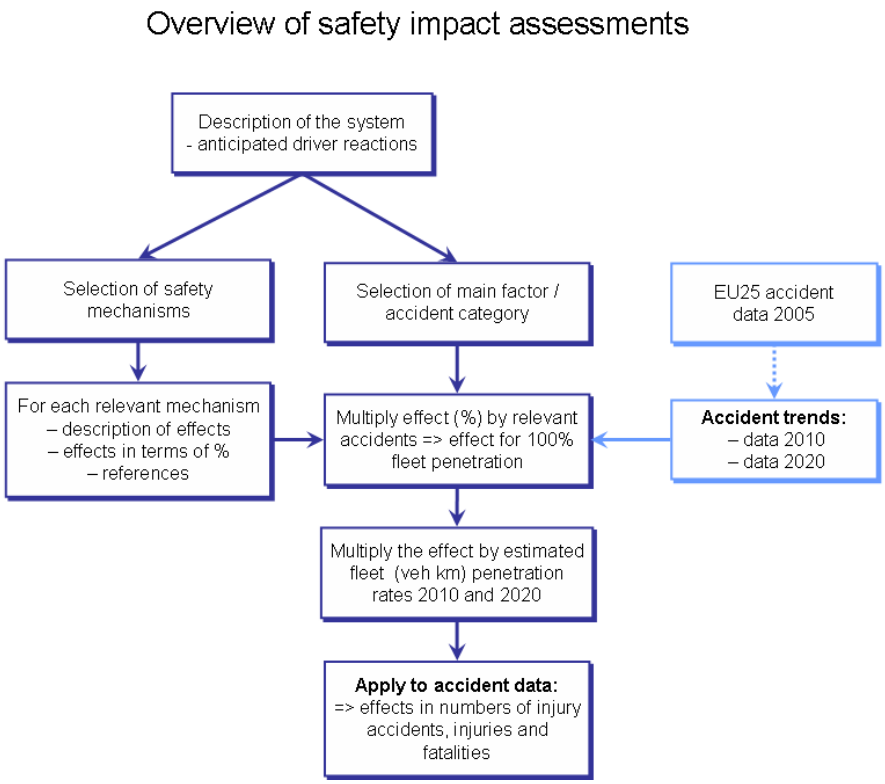

Overview of safety impact assessments
Description of the system
- anticipated driver reactions
Selection of safety mechanisms
Selection of main factor / accident category
EU25 accident data 2005
For each relevant mechanism
– description of effects
– effects in terms of %
– references
Multiply effect (%) by relevant accidents => effect for 100% fleet penetration
Accident trends:
– data 2010
– data 2020
Multiply the effect by estimated fleet (veh km) penetration rates 2010 and 2020
Apply to accident data:
=> effects in numbers of injury accidents, injuries and fatalities

Figure 10.2: An overall schematic picture of the safety impact assessment in eIMPACT

- ***Risk matrix approach***
  The risk matrix approach (RMA) was developed in the euroFOT project. Details about the method can be found in (Van Noort et al., 2012). The RMA is developed for systems that function continuously (e.g. ACC) and that address an isolated accident type (e.g. rear-end collisions). The RMA associates a risk to each data point, by assuming a hypothetical accident scenario, developing from this data point. Separately there is a risk calculation from FOT data. Risks are pre-calculated once, and the application of FOT data is quite simple. This method is a variant of a method developed by NHTSA (W. Najm). The RMA does not rely on video data and is usable without in-depth accident statistics.
- ***Expert judgment***
  This 'method' can be used when quantitative methods (as described above) do not work for some reason, or in addition to it. Expert judgment usually produces qualitative results. Expert judgment can be done in different ways, for example by organising a workshop, or by having experts fill in a questionnaire. Expert judgment should be based on the data that are availble from the FOT (surrogate safety measures e.g. speed, speed variance, headways) but also from previous studies.

The methods mentioned above are not perfect nor applicable in all situations (for all functions, all types of FOTs). In the end one wants to know the relative change (e.g. how much percent of the accidents can be prevented because of a certain system, compared to driving without this system). Other ways to gain insight in changes in safety are for example looking at the frequency of certain events (e.g. hard braking) and speed violations. More on the definition of events can be found in Annex C . From a policy point of view events are better usable than TTC , increase of mean speed, etc.

In the end, independent of the method one chooses, it is important to clearly write down the assumptions that were being made and the consequences of these assumptions. When there is a lot of insecurity in the safety impact assessment, an option is to work with a bandwith and not deliver one fixed number as a result, but a range. A sensitivity analysis can also help with this.

As an example and representing best practice, the Mack FOT puts the goals of the safety analysis as follows:

1. Determine if driving conflict and crash probabilities will be reduced for drivers using the system,
2. Determine if drivers drive more safely using the system,
3. Determine reduction in crashes, injuries, fatalities if all fleets operating in the observed area were equipped with the system,
4. Determine if drivers using the system have less severe crashes than drivers without the system.

The first step collects sensor data from each vehicle within the FOT (e.g. brake force, steering angle). Based on earlier definitions the number of driving conflicts can be determined. Thus, two numbers for the driving conflicts – reflecting the with and the without case – are available to calculate the exposure ratio. This ratio reflects the number of driving conflicts in the with case compared to the without case. To provide an example: given a system which maintains the safe distance to a predecessor vehicle, the number of driving conflicts due to close following will be reduced from 10 conflicts per 1000 km to 5 conflicts per 1000 km. Thus, the exposure ratio equals 0.5 which indicates that driving with the system is safer than without the system. In general, an exposure ratio below 1 indicates a safety benefit.

The benefit of lower exposure to accident risk will likely be modified based on adaptations of individual behaviour due to psychological reasons (second step). Behavioural adaptations can comprise e.g. adapting the following distance, adapting the speed variance, adapting the lane change behaviour (risky cut-ins or changing the lane without signalling it in advance). Examples for such behavioural changes can be found in the ITS safety mechanisms (eIMPACT). In this project, nine mechanisms have been introduced which lead to positive or negative safety effects. In most cases, the motivation for behavioural adaptation is that the driver wants to avoid “public” warnings (noticeable to all passengers) and “education” by the system.

The third step deals with scaling-up from the FOT to a wider area (EU, country, region). This process is subject to the procedure proposed in scaling up.

The last step leads to the prevention ratio. In-depth information on accidents is used to calculate the mitigation effects of using the system. Maybe the system cannot avoid the accident but it can mitigate the accident consequences. This issue has to be considered in determining the effects for casualties. For systems affecting speed, the Power Model can be applied to calculate changes in severity.

Combining steps 2 to 4, it is possible to calculate the prevention ratio. For this ratio the probability of having a crash (casualty) when having a driving conflict in the with case is compared to the same probability in the without case. In the above example the number of driving conflicts in the with case was 5 and 10 in the without case. Let us assume that out of the 5 driving conflicts 1 accident occurs and out of the 10 driving conflicts 3 accidents occur. Thus, the probability of having an accident due to a driving conflict is 0.2 in the with case and 0.3 in the without case. These values reflect the prevention ratios.

10.3.2 Efficiency benefits

Efficiency benefits are typically composed of two effects. They involve:

- Direct mobility effects resulting from a smoother traffic flow, e.g. where the system allows traffic to re-route to avoid current congestion, or improves mean speeds by encouraging safe following behaviour,
- Indirect mobility effects resulting from reduced crashes e.g. reduced delays at incidents and accidents.

Direct mobility effects can play an important role in the socio-economic impact assessment. On the appraisal level, direct mobility effects are reflected in changes of time costs, fuel consumption costs and reliability changes. Because socio-economic impact assessment identifies quite commonly reductions of time costs as a major driver of the results, direct mobility effects are generally worthwhile to explore.

The investigation of direct mobility effects typically involves microscopic traffic flow simulation. A number of models (e.g. ITS Modeller, VISSIM, Paramics, DRACULA) have been applied to assess these impacts. Best practise, including on cross-validation of models, can be found in eIMPACT D4 (Wilmink et al., 2008) and Full Traffic (Technische Universiteit Delft, 2008). Typically, when traffic flow becomes more homogeneous, the standard deviation of the vehicle speed becomes lower. As a result, the average vehicle speed may increase or the infrastructure capacity improves. As a consequence, time costs and vehicle operating costs will decrease.

However, the realisation of those benefits is closely related to the likely market penetration. Mature ICT systems typically can produce such effects, ICT systems in the phase of market introduction typically can not. For internal efficiency it is therefore important to figure out at the beginning of the FOT assessment (when the scope is defined) whether direct mobility effects will be likely to appear or not.

Compared to the direct mobility effects, experience suggests indirect mobility effects are not restricted by conditions of market penetration. They can be realised in any case, as an add-on to the safety benefits. Indirect effects occur when the number — as well as the severity — of crashes is reduced. The benefits result from less congestion, therefore reducing journey times and fuel consumption. Typically, indirect traffic effects add up to about 10 % of the safety benefits.

Given the state of the art in traffic modelling, indirect mobility effects are assessed more frequently than direct mobility effects. Good practise on the appraisal of indirect mobility effects can be found, however, in recent European scale assessment studies (eIMPACT; COWI, 2006) and US American FOT assessments (Batelle Memorial Institute, 2003; Volvo Trucks North America Inc, 2007). Some countries have methods specifically to address these effects (e.g. INCA in the UK).

10.3.3 Environmental benefits

Environmental benefits comprise lower CO_2 and air pollutants emissions. Noise also fits into this category but we would caution that noise should only be analysed where ICT systems are expected to make a significant difference between the two scenarios (with/without the system). CO_2 and pollutants emissions are both speed dependent, with CO_2 emissions directly linked to fuel consumption.

Exhaust emission from road traffic is a complex process to describe. In 5.5.3 a simplified formula is reported for calculation:

Σ(Traffic activity) x (Emission factor)=Total emissions

where Traffic activity data includes: mileage, engine starts and parking. In addition to Traffic activity data one needs data for: the vehicle fleet; road network; meteorological conditions; fuel quality, etc. If the driving pattern is influenced by the traffic situation, such data for the FOT vehicles are directly available. In order to estimate driving pattern changes for all vehicles by traffic situation, microsimulation models could be used. In order to estimate emission factors for these alternative driving patterns there is need for exhaust emission measurements or exhaust emission models on an individual level. The recorded speed traces from the FOT vehicles can also be post-processed through a fuel consumption and emissions model to produce data on environmental effects.

The impact of CO_2 emissions is on a global scale, and is not linked to the particular country or area type where the CO_2 is emitted. The impact does, however, vary according the year in which the reduction (or increase) in emissions takes place — the impact becoming greater further into the future. Actually, mobility effects have impacts on both efficiency and environmental benefits. However, because they are transmitted through the environment, and because they are largely externalities (i.e. their incidence is mostly on individuals other than the emitter) environmental benefits fall into a special category.

10.3.4 Integration of results

FOTs should make sure that there is enough time for the integration of results. At the end of the analysis phase, there are results for traffic efficiency, behaviour, safety, environment, acceptance, etc. This has to result in overall conclusions about a system, where the different impact areas are interwoven. However, this requires sitting together with the experts, and having time to let the results 'sink in'. Often this time is not available.

It is recommended to make a template for the reporting of results early on (for both external and internal reports), so researchers know what is expected from them. Reporting in a clear and systematic way helps with the integration of results and the overall view on the effects of the system.

It is very useful to make room for analysts to add insights discovered during the analysis phase (internal reports). It is a pity if these lessons should be lost as they could prove highly valuable to future FOTs

10.4 Socio-economic Cost Benefit Analysis

The analyst faces choices in setting up and carrying out the analyses. The choices will be influenced by the priorities identified by those setting up the FOT . as well as budget and time constraints. The list below summarises these choices.

Methods:

- The basic choice is CBA, which summarises benefits and costs at a societal level
- Stakeholder perspectives: Makes use of the same input data as the CBA, but considers stakeholder-specific benefits, costs and financial analyses.

Identification of impacts:

- The basic choice is to use the costs incurred and the main expected benefit(s), as identified by use of the impact table.
- Other impacts — both direct and indirect — can also be included, depending on the stakeholder perspective as well as the choices made elsewhere in the project, for example in hypothesis formulation, measurement methods and equipment and modelling capability.
- Willingness-to-Pay evidence, if also collected during the FOT. can be used to supplement the analysis methods above.

Scope of geographical assessment:

- The basic choice is the country level. In this case, the generic data needs (see section 10.4.4) are limited to the country in question.
- EU-level analyses are preferred. These require substantially more general data from individual countries. Extra challenges in execution can be encountered due to differences in definitions or classifications.

10.4.1 System costs

System cost estimation is an element within FOTs which is quite often neglected. System promoters may not see costs as an impact. However from a socio-economic point of view, they are a (negative) part of the impact of systems. Cost estimation should take care of the following aspects:

- **Cost elements to include:** The system costs comprise the costs of in-vehicle, roadside infrastructure equipment and nomadic devices. Besides that, operating and maintenance costs have also to be considered. Examples of good practise for system costs can be found in US American FOT assessments (Freightliner FOT, Mack FOT, Volvo FOT).
- **Relevant size of costs:** CBA applies a resource based view. This means looking at potential savings of productive resources and on the other hand at the resources necessary to achieve this effect. The implication for cost estimation is that only the input of productive resources is relevant and not potential market prices. The convention proposed e.g. by eIMPACT is to use the cost price (the price of the ICT system paid by the manufacturer to its supplier) plus a mark-up which is allowed for in-vehicle implementation. However, the contrary, market prices are relevant for user-centred analyses. Generally, in the face of limited evidence it is useful to apply the "Factor 3" rule of thumb, which means that in the automotive industry market prices for ICT systems differ from the cost prices by a factor of 3.
- **Process of cost estimation:** Typically, cost estimation will be carried out by an expert group comprising of FOT internal staff and external industry experts. To avoid conflicts with confidentiality and the like, it appears sometimes helpful to introduce rough estimations to the group instead of working from blank sheets. Guidance to

rough estimations for investment and OEM costs can be applied from an US-American database on ITS costs and benefits (www.itscosts.its.dot.gov).

10.4.2 Classification of assessment methods

Errore. L'origine riferimento non è stata trovata. gives a classification of socio-economic assessment methods, based on which of the elements are included, and in particular:

- Whether a full set of impacts is addressed — for example, if a significant CO_2 reduction can be anticipated, has it been included;
- Whether the assessment is from the social perspective only, or whether financial and stakeholder analyses are also provided.

The recommendation is that the FOTs should be designed to be as complete as possible, both in terms of impacts and stakeholder views. The assessments in the FOTs reviewed are examples of good practice. However, they differ in the types of analyses carried out, as well as in the scope of the effects examined, with the exception of safety impacts.

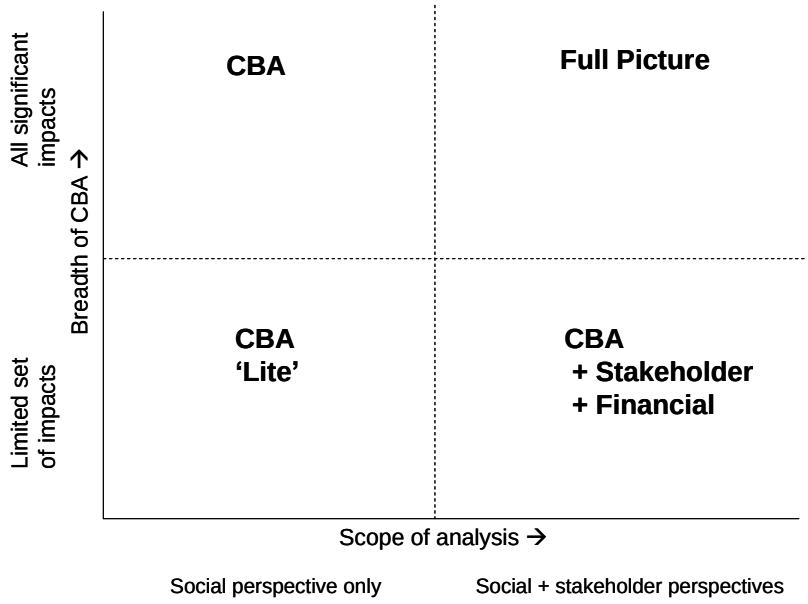


Figure 10.2: Classification of Socio-Economic Assessments

Errore. L'origine riferimento non è stata trovata. highlights another dimension in which assessment methods can be classified, namely whether or not they make use of case-specific Willingness-to-Pay (WTP) evidence. In the design of future FOTs, we recommend that clients and analysts consider WTP studies as a way of getting better evidence on the users' likely demand for the products. WTP can provide uniquely useful evidence on the value of the ICT system to consumers and producers. In absence of this, FOTs can refer to evidence in the literature (market-based). WTP studies will, however, add to the cost and skill set required for FOTs, so the advantages and disadvantages will need to be weighed in each case.

We note that past FOTs have generally relied on market-based values (e.g. the U. S. CAS and Mack FOT), although the U. S. ICC FOT did make use of specific WTP evidence, and as such is a useful reference. Also, we note that most previous assessment guidelines, including eIMPACT, assume that literature-based values will be used. Here, we leave the option open and recommend that clients and analysts decide at the inception phase of the FOT whether or not to go down the WTP route.

10.4.3 How to carry out the assessment

10.4.3.1 Carrying out a CBA

The socio-economic impact assessment of a system within an FOT should be based on a CBA, since it is the most widespread, commonly accepted and practised method for analysing socio-economic impacts. It is clear that CBA accounts for all benefits and all costs on a society level, including benefits and costs to all groups. CBA follow a four-step-process involving framework and preparatory work, measuring impacts, appraising impacts in a common monetary value and confronting the discounted society benefits with the costs of the policy measure. However, this process leaves also some room for shaping the individual steps of the process. We recommend considering the following issues.

1. CBA framework:

Definition of the cases to be compared: Looked at is the with-case (ICT-system equipped) against the without-case (without the system).

Base year and time horizon of the assessment: CBA can be performed for the whole life cycle of the considered system or only for selected target years. This decision depends on information needs.

Geographical scope: Because of data availability the geographical scope should be congruent to existing statistical reporting ICT systems. Reference only to the local area where the FOT takes place is insufficient for this reason and because the results of different FOTs need to be compared. This implies however that the socio-economic impact assessment has to undergo a scaling up procedure before the CBA in order to project the impacts from the FOT itself on to a larger area. The most practical appears to be assessment at the national level (assuming “nationwide deployment”). However, it is even more useful to provide results on a European level. The European perspective is important when the effects of FOTs in different member states should be compared or when policy measures are planned or considered to ensure a European scale deployment (e.g. eCall).

Discount rate: The discount rate ensures that benefits and costs are expressed for a common base year. A discount rate of 3 % (real) is recommended as a default (see 'Other economic parameters').

Deployment scenario: It has to be estimated which share of new vehicles or which share of the total vehicle fleet will be equipped with the system in the target years and over the assessment period as a whole (depends on answer to 'Base year and time horizon' issue

above). For life cycle assessment it is also necessary to estimate the development of the equipment (technical capabilities, costs).

Impact table: The impact table serves as an instrument to expedite identification of impacts. It is aimed to ensure that the FOT team and the group responsible for the socio-economic impact assessment are fully aware of the complete impacts of the system. For efficiency reasons and likely budget constraints (competing FOTs and competing assessment issues within an FOT) it is necessary to concentrate the analysis on the significant impacts — impacts expected to be negligible, or impossible to analyse within the resources available, should be flagged as such in the impact table. Concerning the system, safety is the relevant impact by definition. Direct and indirect mobility impacts and environmental impacts are typically also addressed. System costs will always be relevant.

2. Inputs for impact assessment (including cost estimation)

Impact measurements represent an essential input to the cost-benefit assessment. We would normally expect most of these to feed through from the FOT experiment to the scaling-up procedure (chapter 9) to the CBA inputs. In particular, accident prevention and system costs at the national / EU levels should be delivered this way. Impacts on mobility and environment will typically require additional analysis at the CBA stage (although in a well designed FOT experiment, it may be possible to gather data specifically on any expected sources of benefit, e.g. reduced variability of traffic speeds or reduced fuel consumption; see the TAC Safe Car FOT). The analysis of different FOT assessment has revealed some evidence on best practise for impact measurement. The requirements for CBA can be provided as a sort of output specification. This makes sure that the socio-economic impact assessment will be provided with the appropriate input data for carrying out the assessment. In terms of an output specification the following elements have to be put in place:

- *Accident and traffic performance database*: See section 10.4.4
- *Effectiveness of the system*: These values represent key output of the FOT which have to be provided to the socio-economic impact assessment.
- Procedure for *scaling up the effects* to nationwide/European level
- *Cost estimations*: See section 10.4.1 on system costs.

3. Impact valuation

Methodological base for impact valuation: The general objective of this step is to provide unit values for the physical impacts. Several methods compete in the field of impact appraisal. They can be subdivided in objective approaches (e.g. damage costs, avoidance costs) and subjective approaches (e.g. willingness-to-pay). In European member states, different practises and preferences exist for impact appraisal. A lot of surveying and standardisation efforts have been made by projects like HEATCO (Bickel et al., 2006) to come to common European base. As a general recommendation, it can be stated that unit values for CBA should be based on objective approaches. However, willingness-to-pay information can largely contribute to a higher quality of the assessment when analyses for the users are carried out.

Good practise on unit values: See eIMPACT (Assing et al., 2006), HEATCO and the handbook on external costs of transport.

National or European unit values: This decision corresponds with the geographical scope. Assessment on national level will typically make use of national cost unit rates. For European scale assessment we recommend using the harmonized values contained in HEATCO — note that these are still differentiated by country, but are on a harmonised theoretical basis.

4. Results

Cost-benefit analyses can produce different summary measures of performance. It represents good practise to calculate the Net Present Value (NPV) by summing up all discounted values of benefits (plus sign) and costs (minus sign). Moreover, Benefit-Cost Ratios (BCR) are a very common expression of system profitability which can be calculated by dividing the total benefits by the total costs. It is also practical (see "Base year and time horizon") to calculate “snapshot” BCR for target years. In this case, the costs will be transformed to annual values (using the discount rate) and will be compared to the target year benefits. For FOTs, we recommend the calculation of both figures, NPV and BCR.

For the **social CBA**, we recommend reporting:

- safety benefit (€M);
- other benefits to road users (€M) — mainly time savings, operating cost savings and reliability gains;
- environmental benefits (€M) — including climate change, regional and local air quality effects; noise; and other impacts;
- revenue to operators (€M) — there may be multiple operators, including infrastructure and service operators – each will want to know the impact on themselves (financial), although for the social CBA these revenues may be aggregated;
- costs to operators (€M) — including capital, maintenance and operating costs;
- revenue to automotive OEMs (€M);
- costs to automotive OEMs (€M);
- revenue to government (€M) — including tax revenue changes;
- costs to government (€M) — including investments in R&D and in implementation of ICT systems.

Tabulation of the social CBA is shown in Table 10-1. All entries are at Present Values[16]. A common base year (for prices and discounting) aids comparison across different technology options. RAILPAG (EIB, 2005) has a more detailed breakdown by stakeholders (an 'SE Matrix'), which some analysts may find helpful in presenting the social CBA.

In cases where the public sector expects to contribute to the development or implementation of the system, we recommend also presenting a Benefit:Cost Ratio with respect to public sector support, which HEATCO (Bickel et al., 2006: 41-2) identifies in use by the EC, UK and Switzerland:

$$BCR = \frac{NPV}{PV(PublicSectorSupport)}$$

The calculation of the Benefit-Cost ratio (BCR) is delicate issue in CBA. On one hand the BCR is a very powerful measure, because it applies to the common situation where investment budgets are limited and maximum value for money is required (making best use of a scarce resource). On the other hand the definition of 'costs' (the denominator) can be problematic. As a general rule, the BCR is useful when the denominator is defined in the same way for all options being compared — for example, NPV per unit of central government budget (which would be a BCR of interest to central government). Our recommendation of a BCR with respect to Public Sector Support broadens this to the budget for public expenditure as a whole. This avoids creating an incentive to manipulate the BCR by shifting costs to local and regional government.

In the example shown in Table 10-1, the BCR with respect to public sector support will be 4240/(379-34) = 10.3, which indicates a high social return from each € of public funds contributed.

[16] For more detailed information please refer to FESTA deliverable D2.6.

Table 10-1: Social CBA tabulation

Group	Impact	€M (2008 base)		
		2015	2025	Present Value (Total)
Consumers	Safety benefits	289	299	3715 a
	Other road user benefits	574	606	603 b
	Environmental benefits	63	66	58 c
	...			d
Producers	Revenue	723	780	8520 e
	Costs	248	233	−8311 f
Government	Revenue	3	4	34 g
	Costs	12	14	−379 h

Net Present Value (NPV) = Σa..h	4240

Notes: sign — all negative impacts on the Group affected are shown with a negative sign, thus Costs appear with a negative sign; 2008 base — indicates appraisal at constant general prices using 2008 CPI, and with 2008 as the base year for discounting in the Present Value column.

10.4.3.2 Carrying out a stakeholder analysis

In contrast to CBA, only particular benefits and costs are relevant for particular stakeholders. The reduction of exhaust and CO_2 emissions are not benefits to users, unless they are charged for it (through vehicle-taxes or tolls). The costs of in-vehicle equipment do not represent costs to the government, unless the government agrees to pay for a share of this. The consequence is that ICT systems which are profitable on society level (NPV, BCR) will not be deployed when a relevant stakeholder group is economically impaired. Hence, it is necessary to include stakeholder perspectives in the FOT socio-economic assessment.

Practically, stakeholder analyses make also use of accounting costs and benefits, but on the level of the individual stakeholder group. This implies the following for *users*, but also in general:

- Cost and benefits must be investigated according to their stakeholder relevance. Safety benefits (reduced accident and casualty risks) for instance are relevant to *users* (and to *insurance companies* as well).

- The appraisal of the impacts can be different. *Users* face market prices when considering the investment in a system (see factor 3 rule of thumb). For benefit evaluation the implication is to use market values if available (e.g. fuel consumption: station prices (incl. taxes) instead of net prices). Otherwise, willingness-to-pay approaches have their justification here because they are better suited to reflect individual preferences.

Further adaptations to the CBA approach involve the use of a different discount rate (reflecting private sector interest rates) and the use of a different result measure (fair

market price for a pre-defined annual vehicle mileage or the critical (break-even) mileage for a given market price).

The stakeholder analysis reporting will vary with the analytical methods used. For example, in the TAC Safe Car Project, Monash University used subjective questionnaire methods to investigate users' acceptance of several ICT systems including ISA (see Regan et al., 2006).

Another useful form of stakeholder analysis from the User perspective is Willingness-to-Pay evidence, as shown in eIMPACT D3 (Assing et al., 2006 p. 119).

For the *vehicle* OEMs and both *infrastructure and service operators*:

- Where they are commercial bodies, a *financial analysis* will provide the most important stakeholder information;
- Where they are *public sector* agencies, a *financial analysis* may need to be combined with an assessment against their public service objectives – however, in some cases the overall social CBA will serve this purpose, depending on the approach taken by the agencies involved.

10.4.3.3 Carrying out a financial analysis

The internal rate of return (IRR) of a project is the interest rate that will generate an NPV of zero. In an equation, this is:

$$\sum_{t=0}^{T} \frac{B_t - C_t}{(1+i_{IRR})^t} = 0$$

where IRR is internal rate of return.

The stakeholder for whom the IRR is calculated compares the IRR with a target rate. This target rate depends for each stakeholder. For public authorities as a stakeholder the target rate will be less than for private investors as stakeholders.

In any case, a calculated BCR or IRR should be accompanied by an NPV. We recommend that financial IRRs are reported for all FOTs.

The IRR concept can be modified for comparison reasons. For his approach, the cash flow streams are subtracted. With the new cash flows the modified IRR is calculated. If the IRR is above the trigger rate, the project with the larger cash flow is the better project.

Of key interest will be the IRR from the point of view of *specific stakeholders* (or stakeholder types). The IRR for vehicle OEMs will influence their decision about investing in the technology. Similarly, the IRR for infrastructure operators and service operators will influence their decisions – particularly where these are commercial operations.

Hence the key information will be in the form:

IRR_{OEMs} = … %

$IRR_{RoadAuthorities}$ = … %

Further IRRs should be reported where there are other stakeholders with a commercial interest, for whom significant impacts are expected. Tables such as those used by WebTAG (DfT, 2005) also provide a useful series of snapshots of the financial impact. In this case, in order to be meaningful the tables should relate to specific stakeholders or stakeholder types, e.g. vehicle OEMs or road authorities.

The financial results can be taken a stage further by reporting the breakeven point in terms of sales or market penetration, or the target price, down to which the system must be engineered in order to achieve financial viability. Graphical presentations may be useful in these cases.

10.4.4 Data needs

The data needed to carry out a socio-economic assessment for an FOT are extensive, and fall into two broad categories:

- FOT-specific data which will be gathered during the FOT itself
- Generic data, which play a role in:
    - Scaling up the results from the experimental situation of the FOT to the national or EU level
    - Reaching a socio-economic assessment, based on the FOT data scaled-up to National or EU level.

The following section outlines the FOT-specific and generic data likely to be needed. Thereafter recommendations on ensuring data quality and validity are given. Management of the data for socio-economic assessment is after this.

10.4.4.1 FOT-Specific Data

The key items of FOT-specific data likely to be needed are:

Accident rates (or risks) with and without the ICT system in place for the FOT sample. These will need to be differentiated by all the key drivers of accident rates (risks) in the FOT sample (e.g. road type; driver type; traffic conditions) so that accurate extrapolations can be made to the whole network. Accident rates (or risks) will be needed with and without the ICT system in place for the FOT sample. These may need to be derived from data on unsafe behaviours if the sample is too small to contain a significant number of actual accidents, although this is likely to be done as part of the performance indicators in any case. See section 10.3.1 on Safety Benefits.

One approach to estimating the impact on accident rates uses the effectiveness rate (% of relevant crash type avoided) as in the Collision Avoidance Systems (CAS) Benefits Study (NHTSA Benefits Working Group, 1996).

A more sophisticated approach can produce data on accident severity as well as accident rates. Since accident severity is determined by the severity of the most serious casualty only, a complementary item of data would be any expected change in the number of casualties per accident. Regan et al. (2006) measured time spent buckled-up and time before buckling-up to produce injury severity estimates.

Examples of how data is produced for accident severity and accident rates can be found in Regan et al. (2006), UMTRI et al. (2006), Volvo Trucks North America et al. (2007) and USDoT (1999).

Whichever approach is used to estimate accident rates and accident severity, the analysis will need to take account of any options in the implementation path. For example, in the Freightliner FOT study (Batelle Memorial Institute, 2003) there were four possible deployment groups (Hazardous Materials tankers; all tankers; tractor trailers; all large trucks) — input data will be required for each of these options.

Multiple scenarios may also be needed to enable sensitivity testing. That is, where there is uncertainty over accident rates/severity or other key variables, this can be handled through 'what if' scenarios based on combinations of the possible outcomes (Batelle Memorial Institute, 2003).

There may also be some value in having spatially differentiated data, and being able to link behaviour to traffic conditions, e.g. urban / nonurban and traffic congestions (Technische Universiteit Delft; 2008, Volvo Trucks North America et al., 2007).

Market penetration forecasts: In the literature, SEiSS (VDI/VDE-IT, 2005 and Baum et al., 2006) gives particular attention to market penetration.

Usage, reliability and compliance: Although the CAS Benefits Study (NHTSA, 1996) made assumptions about usage, reliability and compliance rather than gathering data, it did draw attention to these important factors in the out-turn effectiveness of ICT systems. Usage refers to the percentage of drivers (or of driving time) for which ICT systems installed on the vehicle will be switched-on and active. Reliability refers to the likelihood that ICT systems will operate without failure, technically. Compliance refers to the percentage of occasions on which the driver's behaviour complies with warning or indication provided by the system.

Attitudinal and acceptance data: Many FOTs gathered attitudinal and acceptance data (Regan et al., 2006; UMTRI et al., 2006; USDoT, 1999; NHTSA, 2006; and Volvo Trucks North America, 2007).

Costs of the ICT systems: See section 10.4.1. In some FOTs, data has been gathered which inputs directly into the maintenance and operating cost calculations (Volvo Truck FOT). Where the assessment period is longer than the expected service life of the equipment, replacement costs should be included (e.g. in the Freightliner FOT one round of replacement was included since the service life was 10 years and the assessment period 20 years; Battelle Memorial Institute, 2003).

10.4.4.2 Generic Data

The key items of generic data likely to be needed are:

National and EU-level network, fleet and traffic data, which are used in scaling-up the findings from the FOT to the level of political interest: The International Road Traffic and Accident Database, IRTAD (ITF, 2008) contains traffic data for the EU27. This includes vehicle kilometres on the total road network, vehicle kilometres on motorways, and vehicle kilometres on urban roads. Vehicle kilometres on rural roads can be derived; some data are missing.

ProgTrans European Transport Report (latest version: 2007/08) can be used as a valid source for forecasts. It contains vehicle stock and vehicle kilometres for 1) cars, 2) buses and coaches and 3) goods vehicles. Generally, the report covers past, present (incl. short-term forecasts for the next years) and future (longer term for selected target years). In the 2007/08 report the following years are covered: 1995, 2000, 2005, 2006, 2007, 2008, 2015 and 2020. Geographically, they cover EU-27 by member state plus some more (Norway, Switzerland, Croatia, Turkey, Belarus, Russia, Ukraine plus China, Japan and the USA).

Accident data (accidents, fatalities, severe and slight injuries) for base scenario: National databases are available. At the EU-level, the collection and compilation fo accident data as a basis for the safety impact assessment is a challenge, especially when specific target accidents are going to be explored. Several EU-projects are dedicated to harmonising accident databases, See TRACE (www.trace-project.org) or SafetyNet (http://erso.swov.nl/safetynet/content/safetynet.htm) for more information. Forecasts of road safety are needed. An example can be found in eIMPACT "Impact Assessment of Intelligent Vehicle Safety Ssytems", in which road safety predictions for 2010 and 2020 for the EU-25 are presented.

More detailed network specifications (e.g. infrastructure equipment) may be required for some systems: the presence/absence of beacons, signalisation. Basic figures (e.g. share of Trans-European Road Network equipped with dynamic traffic management) are available from the eSafety Forum Implementation Roadmap Working Group.

Speed-flow relationships or network models, which allow journey times and costs to be derived from changes in flows: Although these are strictly much more than just 'data', it is worth highlighting the key role they play in socio-economic assessment of transport ICT systems. Many of the effects of new ICT systems will be mediated through changes in traffic flow on the network – for example, advanced warning ICT systems allow drivers to change route to avoid hazards, but the net effect on travel times and costs is dependent not only on the behaviour of the individual, but on the behaviour of large numbers of individuals and the interaction with the limited capacity of the network. Hence, at the very least, knowledge of speed-flow relationships is needed to understand the consequences of shifting traffic across the network.

HCM (2000) and FGSV (2001) are sources of speed flow relationships. Network models or strategic transport models incorporate this data and have much wider functionality. The fact that these models are very expensive to develop and maintain means that they tend not to

be developed for one socio-economic assessment in isolation. Instead, part of the socio-economic assessment process is usually to identify models already existing which can provide the necessary functionality.

Evidence on accident costs, used to measure the benefits of accident reduction and changes in accident severity: The HEATCO project (Bickel et al., 2006) was designed specifically to provide harmonised cost estimates for socio-economic assessment in Europe. We recommend that the HEATCO accident cost values are used in the FOTs, and we provide one additional piece of evidence to fill a gap in HEATCO which is a generic dataset on the costs of 'damage only' accidents.

Two of the main issues in this field are:

- An apparent inconsistency between 'willingness-to-pay' (WTP) methods and 'cost of damage' or 'human capital' methods as a basis for values — empirically, WTP methods can produce significantly higher values for fatalities in particular (see Assing et al., 2006: Table 16);
- Double counting of casualties' lost future consumption, which is included in both lost future output and WTP to reduce accident risk.

HEATCO addresses these issues by specifying a common framework in which the different elements of accident costs measured by each method can be reconciled. For example, 'human capital' methods do not capture people's full valuation of safety risk, whilst WTP-based values do not capture the external resource costs of accidents (e.g. healthcare costs borne by the state) but often do double-count lost future consumption, as already noted. The HEATCO framework includes:

- property damage;
- medical costs;
- administration costs;
- lost output;
- welfare losses due to casualty reduction.

As a result, the HEATCO values for fatalities are neither as high as the US NHTSA's willingness-to-pay values cited by Assing et al., nor are they as low as the NHTSA's cost-of-damage values. They are broadly in line with 'best practice' European values used in cost-benefit analysis, and the differences can generally be understood by examining the differences in the underlying measurement methods.

Other important functions of the HEATCO values are to provide:

- A common unit of account in the face of taxes and subsidies – HEATCO values are provided at the factor cost unit of account (Bickel et al., 2006: 52);
- A common price base year;

- A common currency, €, for European-level assessments.

In HEATCO accident values are listed in a table, expressed as values per casualty saved. These values do include the full set of accident costs, per casualty.

To apply these values, analysts will require further data:

1. Forecasts of accidents with and without the technology in place – based on the FOT findings and the results of the scaling up process.

2. If these forecasts do not address unreported accidents, then factors for the number of unreported accidents given the number of reported accidents can be found in HEATCO (Bickel et al., 2006, Table 5.1).

3. Growth in the values over time – an elasticity of 1.0 with respect to GDP per capita, thus a 2.0 % annual increase in GDP per capita would imply a 2.0 % annual increase in the values of accident reduction.

4. Damage only accident values. As Baum et al (2007) shows, savings in damage only accidents can make up a large proportion of the benefits from ICT safety systems. Damage only accident costs may be approximated at 17 % of the cost of a Slight Casualty (Nellthorp et al., 1998).

National level assessments may wish to take advantage of the most recent safety valuation evidence at the national level. For multi-national assessments it will be important to ensure that any national evidence is checked for consistency across boundaries, and conversions made if necessary (e.g. in terms of base year, unit of account, cost elements included, measurement methodology, etc).

For EU-level assessments, consistency across countries and comparability between assessments will be important, which makes the use of a harmonised set of values (as above) more attractive. If the harmonised values are found not to provide the detail which the analyst wants — e.g. if differentiated accident costs by road type or user type are expected to be a key requirement for a particular assessment — then it may be appropriate to vary the values above, based on more detailed information (for example, the accident cost data included in national level assessment guidelines).

Evidence on values of time savings and vehicle operating cost savings, used to measure the benefits of changes in traffic flow: Values of travel time savings will be needed to assess the benefits of improved traffic flow due to the ICT systems. HEATCO Tables 4.6-4.8 provide suitable values for working and non-working passenger trips, and for freight transport (Bickel et al., 2006: 73-75). These values increase with GDP per capita, at an inter-temporal elasticity of 0.7.

Sometimes there will be an impact on reliability, not only expected (mean) travel times, and in these cases we recommend using the reliability ratios set out in HEATCO Table 4.3 to value changes in the standard deviation of journey time.

Vehicle operating cost savings are also likely to arise from changes in traffic flows. The traffic models used to predict traffic flow responses to ICT systems will typically be capable of predicting changes in Vehicle Operating Costs, and the fine network detail in these models usually makes it more logical to calculate these cost savings within the model, rather than attempting to do so based on model outputs. As a result, standard values are not offered for these impacts by HEATCO (Bickel et al., 2006: 135-140).

Emissions factors and values for the damage caused by emissions of greenhouse gases, air pollutants and noise: HEATCO provides values for both sources of emissions (Bickel et al., 2006, Tables 6.2, 6.4). Values for particulate (smoke) emissions are differentiated between urban and non-urban locations, due to their much localised impact pathway. Other air pollutants are valued uniformly at country level. HEATCO provides a shadow price of CO_2 by year of emissions (Bickel et al., 2006, Table 6.12) which should be applied to all forecast changes. The impact of noise changes may be quantified using the HEATCO values for road, rail and aircraft noise in each member state (Bickel et al., 2006, Table 6.9).

Other economic parameters such as the social discount rate: Discount rates are required for socio-economic assessment. In line with HEATCO, we recommend using a risk-free social time preference rated for the countries to which the assessment would apply. If a default discount rate at the EU level is required, we would recommend using 3 % per annum (real). GDP growth data for the members of the EU27 (required for updating values of accidents, etc, over time) is available from Eurostat.

10.4.4.3 Data quality and validity

The EC ROSEBUD project provided the following guidance as part of a "professional code for analysts" (BASt et al., 2005, p. 46):

"Data has to be attributed correctly to its sources, especially when different data sources like national or international accident databases or in-depth databases are used. Where and how estimations were made to fill data gaps needs to be documented. Regression models should be used to generate future time series; trend extrapolations can replace them where available data are insufficient for regressions".

In addition, we would recommend that:

- The principles of statistics apply — statistical tests should be used wherever possible to determine if hypotheses about ICT system impacts are supported by the FOT evidence, and sample sizes should be chosen to obtain statistically significant results;
- When scaling-up from the FOT to the national or EU27 level, a methodical approach based on the key drivers of safety/other significant outcomes identified in the FOT should be used (cross reference);
- Confidence intervals as well as mean data should be recorded for key variables – note that confidence intervals are given in HEATCO for the various economic parameters recommended;

- We have noted the need to recognise the uncertainty in the data using sensitivity analysis – if analysts wish to take a more advanced approach and use Monte Carlo simulation or related techniques (for example to derive a probability distribution on NPV or BCR) that would be welcome as it simplifies the outputs seen by the decision-makers, although it does place an additional burden on the analysts;

- Known problems with the data should be acknowledged and acted upon, e.g. UMTRI et al. (2006) excluded a proportion of drivers whose trials were invalidated (in that case 9 out of 87 drivers), and some trips by the remaining drivers. Well-known problems with the omission of unreported accidents from data have prompted Bickel et al. (2006, Table 5.1) to provide adjustment factors for different accident severities and types.

Record keeping and data storage are important. This includes qualitative/subjective data, and evidence gathered during deliberative studies (e.g. UMTRI et al. (2006) ensured that focus group evidence was captured on video and by a court stenographer).

Finally, the US NHTSA observes that "the validity of any experimental test results depends on the experimental condition effects that were placed on the drivers" (NHTSA, 1996: p36). Care is needed, therefore, when extrapolating data from short-term experiments to long-term term adjustments in behaviour and demand for ICT systems — e.g. the CAS Benefits Study proposed that "a better estimation of the safety benefits … can be achieved as more relevant test data are gathered especially from long-term, large-fleet field operational tests" (p.C-8).

11. References

Aarts, L. and van Schagen, I. (2006). Driving speed and the risk of road crashes: a review. Accident Analysis and Prevention, 38, pp. 215-224.

Abhishek, V. (2008). European Telematics & ITS. A progress report on V2X communication in Europe.

Ajzen, I., & Fishbein, M. (1980). Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ: Prentice-Hall.

Armstrong, M. (1999). A handbook of Human Resource Management Practice. London, Kogan Page.

Antin, J., Lee, S., Hankey, J., and Dingus, T. (2011). Design of the In-Vehicle Driving Behavior and Crash Risk Study. In Support of the SHRP 2 Naturalistic Driving Study. SHRP2 Report S2-S05-RR-1, ISBN: 978-0-309-12895-7, National Academy of Sciences.

Assing, K., Baum, H., Bühne, J.-A., Geißler, T., Grawenhoff, S., Peters, H., Schulz, W. H. and Westerkamp, U. (2006). Methodological framework and database for socio-economic evaluation of Intelligent Vehicle Safety Systems. Deliverable D3, eIMPACT: Socio-economic impact assessment of stand-alone and co-operative Intelligent Vehicle Safety Systems (IVSS) in Europe, EU 6th Framework Programme Contract no.: 027421. TNO, Delft. www.eimpact.eu

BASt, DITS, KuSS, TRL, UOC and CDV (2005). Road safety and Environmental Benefit-Cost and Cost-Effectiveness Analysis for Use in Decision Making, Deliverable D7. Funded by EC DG-TREN, under project GTC2/2000/33020. http://partnet.vtt.fi/rosebud/index.html

Batelle Memorial Institute (2003). Evaluation of the Freightliner Intelligent Vehicle Initiative Field Operational Test, Final Report to the US Department of Transportation, Project DTFH61-96-C-00077 Workorder 7718. USDoT, Washington DC. http://www.itsdocs.fhwa.dot.gov//JPODOCS/REPTS_TE//13871.html

Baum, H., Geißler, T., Grawenhoff, S., Schulz, W. H. (2006). Cost-Benefit Analyses of Intelligent Vehicle Safety Systems – some empirical case studies. Zeitschrift für Verkehrswissenschaft. 77(3), pp. 226-254.

Baum, H., Grawenhoff, S. and Geißler, T. (2007). Cost-Benefit-Analysis of the Electronic Stability Program (ESP). Summary Report. Institute for Transport Economics, University of Cologne.

Besseling, H. and van Boxtel, A. (2001). Intelligent Speed Adaptation — Results of the Dutch ISA Tilburg trial. Ministry of Transport, Public Works, and Water management, AVV Transport Research Center.

Bickel, P., Friedrich, R., Burgess, A., Fagiani, P., Hunt, A., De Jong, G., Laird, J., Lieb, C., Lindberg, G., Mackie, P., Navrud, S., Odgaard, T., Ricci, A., Shires, J., Tavasszy, L. (2006). Developing Harmonised European Approaches for Transport Costing and Project Assessment, HEATCO Deliverable 5. Germany. http://heatco.ier.uni-stuttgart.de/

Boyle, L. N., Lee, J. D., Neyens, D. M., McGehee, D. V., Hallmark, S., & Ward, N. J.

(2009). SHRP2 S02 Integration of analysis methods and development of analysis plan. Phase 1 Report. University of Iowa, Iowa City, IA.

Carsten, O, Fowkes, M., Lai, F., Chorlton, K., Jamson, S., Tate, F. and Simpkin, B. (2008). ISA-UK: Final Report. Institute for Transport Studies, University of Leeds.

Castells M. (2001) The Internet Galaxy: Reflections on the Internet, Business and Society. Oxford University Press, 2001

Cochran W. G. (1977). Sampling techniques. 3rd edition. Wiley & Sons, New York.

COWI, ECN, Ernst & Young Europe, ECORYS (2006). Cost Benefit assessment and prioritisation of vehicle safety technologies. Final Report.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., and Knipling, R. R. The 100-Car Naturalistic Driving Study, Phase II — Results of the 100-Car Field Experiment Virginia Tech Transportation Institute, Blacksburg, VA, NHTSA DOT HS 810 593.

Draft ETSI EN 302 637-2 V1.3.0 (2013-08): Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service

Draft ETSI EN 302 637-3 Intelligent Transport Systems (ITS); Vehicular Communications;Basic Set of Applications; Part 3: Specifications of Decentralized Environmental Notification Basic Service

Draskóczy, M., Carsten, O. and Kulmala, R. (1998). Road safety guidelines. Deliverable B5.2 of CODE project (TR1103). Atkins Wootton Jeffreys, Birmingham, UK. Available at: http://cordis.europa.eu/telematics/tap_transport/library/code-b5-2.html.

Ehrlich, J. et al. (2003). LAVIA, The French ISA project: main issues and first technical results, the French ISA project. Proceedings of the 10th World Congress on Intelligent Transport Systems. Madrid, Spain, 16-19 November 2003.

Elvik, R., Christensen, P. and Amundsen, A. (2004). Speed and road accidents: an evaluation of the Power Model. TOI Research Report 740/2004, Institute of Transport Economics, Oslo.

Ervin, R. D., Sayer, J., LeBlanc, D., Bogard, S., Mefford, M., Hagan, M., Bareket, Z., Winkler, C. (2005). Automotive collision avoidance system field operational test methodology and results, Volume 1: Technical Report. UMTRI-2005-7-1; University of Michigan, Ann Arbor, Transportation Research Institute.

ETSI EN 302 663 V1.2.1 (2013-07) Access layer specification for Intelligent Transport Systems operating in the 5 GHz frequency band

ESOP 99 Commission of the European Communities (1999). Commission Recommendation of 21 December 1999 on safe and efficient in-vehicle information and communication systems:

A European statement of principles on human machine interface. Doc. no. C(1999) 4786, 2000/53/EC.

Everitt, B. and Dunn, F. (2000). Applied multivariate data analysis. Arnold Publication.Faber, F., Jonkers, E., Ljung Aust, M., Benmimoun, M., Regan, M., Jamson, S., Dobberstein, J. (2011). EuroFOT Deliverable 6.2, Analysis methods for user related aspects and impact assessment on traffic safety, traffic efficiency and environment. June 30 2011.

Fancher, P., Ervin, R., Sayer, J., Hagan, M., Bogard, S., Bareket, Z., Mefford, M. & Haugen, J. (1998). Intelligent Cruise Control Field Operational Test. Final Report UMTRI-98-17; University of Michigan, Ann Arbor, Transportation Research Institute.

Featherman MS, Pavlou PA (2003). Predicting e-Services Adoption: A Perceived Risk Facets Perspective. Int. J. Hum. Comput. Stud. 59(1)

FGSV (2001). Forschungsgruppe für Strassen- und Verkehrswesen. Handbuch für die Bemessung von Straßenverkehrsanlagen. FGSV, Köln. http://www.fgsv.de/

Fitch, G. M., H. A. Rakha, et al. (2008). Safety Benefit Evaluation of a Forward Collision Warning System, US DOT, NHTSA, Report No DOT HS 810 910, Washington, D.C.

FOT-Net 2 D3.2: Working Group Additions to FOT Methodology

Gibbons, J. D. and Chakraborti, S. (2003). Nonparametric statistical inference. 4th edition. CRC.

Goldstein, H. (2003). Multilevel statistical models. 3rd edition, Arnold, London.

Grudin, J. (1992) Utility and usability: research issues and development contexts. Interacting with Computers, Volume 4, Issue 2, August 1992, Pages 209-217

Hanowski, R. J., M. Blanco, et al. (2008). The Drowsy Driver Warning System Field Operational Test: Data Collection Methods, Virginia Tech Transportation Institute.

HCM (2000). Highway Capacity Manual. Transportation Research Board, Washington D.C.

Hjälmdahl, M., Várhelyi, A. (2003). Speed regulation by in-car active accelerator pedal effects on driver behaviour. Transportation Research Part F. Vol. 7, Issue 2, pp. 77-94.

ITF (2008). The International Road Traffic and Accident Database (IRTAD). www.irtad.net

Jacobson, I., Bylund, S., Jonsson, P., and Ehneboom, S. (1995), “Modeling with Use Cases: Using contracts and use cases to build plugable architectures”. Journal of Object Oriented Programming, Vol. 8, No. 2, pp. 18-24.

Jamson, S., Chorlton, K, Gelau, G., Schindhelm, R., Johansson, E., Karlsson, A-S, Metz, B,, Tadei, R., Benmimoun, M., Val, C., Regan, M., Wilschut, E. Brouwer, R. (2009). Experimental Procedures. Deliverable 4.2 of EuroFOT. Ford Research & Advanced Engineering Europe, Aachen, Germany.

Kato, K. et al. (1999). Plan for verification experiments of advanced cruise-assist highway system (AHS) in Japan. Proceedings of 6th World Congress on Intelligent Transport Systems (ITS), held Toronto, Canada; http://www.ahsra.or.jp/eng/d01e/index.html

Lassarre, S. and Saad, F. (2006). Présentation générale du dispositif expérimental: justification des choix. In: Carnet de route du LAVIA. Limiteur s'adaptant à la vitesse autorisée. Paris: Actes du colloque LAVIA, pp. 11-17.

Lassarre, S. and Romon, S. (2006). Utilization du LAVIA et influence sur les vitesses pratiquées en vue de l'évaluation de l'utilité. In: Carnet de route du LAVIA. Limiteur s'adaptant à la vitesse autorisée. Paris: Actes du colloque LAVIA, pp. 53-60.

Lebart, L., Morineau, A. and Piron, M. (1997). Statistique exploratoire multidimensionnelle. Dunod, Paris.

Lecoutre, J. P. and Tassi, P. (1987). Statistique non paramétrique et robustesse. Econometrica, Paris.

Li, B. (2004). The latest development of intelligent highway system in China. Proceedings of the 11th World Congress on ITS, Nagoya, Japan.

D. Mans, E. Jonkers, I. Giannelos, D. Palanciuc (2013). Scaling up methodology for CO_2 emissions of ICT applications in traffic and transport in Europe. ITS Europe congress in Dublin, 2013.

Michon, J. A. (1985). A critical view of driver behaviour models: what do we know, what should we do? In: L. Evans and R. C. Schwing (Eds.). Human Behaviour and Traffic Safety, pp. 485-524. New York: Plenum Press.

Najm, W. G., Stearns, M. D., Howarth, H., Koopman J. and Hitz, J. (2006). Evaluation of an automotive rear-end collision avoidance system. DOT HS 810 569. DOT VNTSC-NHTSA-06-01. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Nellthorp, J., Mackie, P.J. and Bristow, A.L. (1998) Measurement and Valuation of the Impacts of Transport Initiatives, Deliverable D9, EUNET Project, EI Fourth Framework RTD Programme. ITS, University of Leeds, Leeds, UK.

NHTSA Benefits Working Group (1996). Preliminary Assessment of Crash Avoidance Systems Benefits. NHTSA, Washington D.C. http://www.trb.org/news/blurb_detail.asp?id=1166

Nielsen, J. (1993). Usability engineering. Boston, MA: Morgan Kaufmann Publishers.

Nilsson, G. (2004). Traffic safety dimensions and the power model to describe the effect of speed on safety. Bulletin 221. Lund Institute of Technology, Lund University. http://www.lub.lu.se/luft/diss/tec_733/tec_733.pdf

Oguri, Y. (2007). Viewing traffic safety issues from victims' standpoints – developing wider social acceptance of safe speed initiatives. ICTCT Extra Workshop, Beijing.

Olson, R. L., Hanowski, R. J., Hickman, J. S., and Bocanegra, J. (2009). Driver Distraction in Commercial Vehicle Operations. FMCSA DTMC75-07-D-00006

Orban, J., Hadden, J., Stark, G. and Brown, V. (2006). Evaluation of the mack intelligent vehicle initiative Field Operational Test; Final Report. FMCSA-06-016. Batelle, Columbus, Ohio.

Özkan, T. and Lajunen, T. (2005). Multidimensional traffic locus of control scale (T-LOC): factor structure and relationship to risky driving. Personality and Individual Differences,38, 533-545

Page, J. (2004). A final technical report on the Belgian intelligent speed adaptation (ISA) trial. 11th World Congress on ITS in Nagoya, Japan. IBSR-Report V05/04 available in French and Dutch.

Polit, D.F., Beck, C.T. and Hungler, B.P. (2001), Essentials of Nursing Research: Methods, Appraisal and Utilization. 5th Ed., Philadelphia: Lippincott Williams & Wilkins

ProgTrans (2004). European Transport Report. Basel.

Regan, M. A., Lee, J. D., and Young, K. L. (2008). Driver distraction: Theory, effects and mitigation. Florida, USA: CRC Press. (In Press – available October 2008).

Regan, M., Triggs, T., Young, K., Tomasevic, N., Mitsopoulos, E., Stephan, K., and Tingvall, C. (2006). On-road evaluation of intelligent speed adaptation, following distance warning and seatbelt reminder systems: Final Results of the Australian TAC SafeCar Project. Volume 1: Report. Monash University Accident Research Centre Report 253. MUARC: Melbourne, Australia. http://www.monash.edu.au/muarc/reports/muarc253.html

Saad, F. (1997). Contribution of observation and verbal report techniques to an analysis of road situations and drivers' activity. In: T. Rothengatter and Carbonell Vaya (Eds.). Traffic and Transport Psychology, Theory and Application, pp. 183-192. Pergamon.

Saad, F. (2006). Some critical issues when studying behavioural adaptations to new driver support systems. Cognition, Technology and Work, 8, pp. 175-181.

Saad, F. and Dionisio, C. (2007). Pre-evaluation of the “Mandatory Active” LAVIA: assessment of usability, utility and acceptance. In: Proceedings of the 14th World Congress and Exhibition on Intelligent Transport Systems and Services. 8-12 October 2007, Beijing, Paper 2257.

Sagberg, F, et al (2011). Recommendations for a large-scale European naturalistic driving observation study. PROLOGUE Deliverable D4.1. Grant agreement no. 233597.

Sanghoon, B. (1998). Evaluation of ITS Field Operational Test in Kwachon Korea. The 5th ITS World Congress, Seoul, Korea, October 12-16, 1998.

Sanghoon, B. (1998): Issues and lessons learned from ITS Field Operational Test in Korea. The 1st Asia Pacific Conference on Transportation and the Environment, Singapore, May 13-15.

Särndahl, C.-E., Swensson, B. and Wretman, J. (1992). Model assisted survey sampling. Springer Verlag.

Shakel, B. and Richardson,S.J. (1991) Human factors for informatics usability. Cambridge University Press

Taylor H. and Karlin S. (1994). An introduction to stochastic modeling. Academic Press.

Technische Universiteit Delft, Faculteit Civiele Techniek en Geowetenschappen, Afdeling Transport & Planning (2008). Full Traffic – WP Dataloggers & WP Verkeersimpact. TU Delft, Netherlands.

University of Michigan Transportation Research Institute (UMTRI), Visteon Corporation and AssistWare Technology Inc. (2006). Road Departure Crash Warning System Field Operational Test: Methodology and Results. Volume 1: Technical Report and Volume 2: Appendices. NHTSA, Washington D.C.
http://deepblue.lib.umich.edu/bitstream/2027.42/49242/1/99788.pdf and
http://deepblue.lib.umich.edu/bitstream/2027.42/49242/1/99789.pdf

U.S. Department of Transportation (2005). Volvo Trucks Field Operational Test: Evaluation of Advanced Safety Systems for heavy truck tractors. Final Report. DTFH61-99-X-00102. Washington, DC: US DOT Federal Highway Administration.

U.S. Department of Transportation, NHTSA (2006). Automotive Collision Avoidance System Field Operational Test Report: Methodology and Results.
http://deepblue.lib.umich.edu/bitstream/2027.42/49539/1/99798.pdf

Van Noort, M., Faber, F., Bakri, T. (2012). EuroFOT Safety impact assessment method and results. 91st Annual Meeting of the Transportation Research Board, Washington D.C., January 2012.

VDI/VDE-IT and ITE University of Cologne (2005). Exploratory Study on the potential socio-economic impact of the introduction of Intelligent Safety Systems in Road Vehicles (SEiSS). Final Report to the European Commission DG Information Society and Media.
www.esafetysupport.org/en/esafety_activities/related_studies_and_reports

Victor, T., et al (2010). Sweden-Michigan Naturalistic Field Operational Test (SeMiFOT) Phase 1: Final Report. SAFER Report 2010:02. Göteborg, Sweden.

Volvo Trucks North America Inc (2007). Volvo Trucks Field Operational Test: Evaluation of Advanced Safety Systems for heavy truck tractors. Final Report to US Department of Transportation, Federal Highway Administration, Cooperative Agreement No. DTFH61-99-X-00102. USDoT, Washington DC. http://wiki.fot-net.eu/index.php?title=Volvo_Intelligent_Vehicle_Initiative_Field_Operational_Test
Wannacott, T. and Wannacott, R. (1990). Introductory statistics for business and economics. 4th edition. John Wiley & Sons.

Wasserman L. (2007). All of nonparametric statistics. Springer.

Wilmink, I., Janssen, W. Jonkers, E., Malone, K., van Noort, M., Klunder, G., Rämä, P., Sihvola, N., Kulmala, R., Schirokoff, A., Lind, G., Benz, T., Peters, H., Schonebeck, S. (2008). Socio-economic Impact Assessment of Stand-Alone and Co-operative Intelligent Vehicle Safety Systems (IVSS) in Europe. Deliverable D4. eIMPACT: Socio-economic Impact Assessment of stand-alone and co-operative Intelligent Vehicle Safety Systems (IVSS) in Europe, EU 6th Framework Programme. TNO, Delft, 2008. http://www.eimpact.eu

List of Annexes

Annex A Legal and ethical issues in the execution of FOTs – Worked Example

Annex B FOT Implementation Plan (FOTIP)

Annex C Defining crash relevant events in NDS/FOT studies formulation

FESTA Handbook

Version 5

Annex A Legal and ethical issues in the execution of FOTs – Worked Example

---

Revised by FOT-NET

(Field Operational Tests Networking and Methodology Promotion)

28 FEBRUARY 2014

Annex A Legal and ethical issues in the execution of FOTs – Worked Example

The aim of this Annex is to sensitise the reader to the legal issues that will prove to be relevant in planning and carrying out a Field Operational Test (FOT). Due to the fact that the details of future field tests cannot be foreseen, all obviously relevant legal areas are covered; necessarily giving abstract information devoid of any warranty as far as completeness and accuracy for the concrete test arrangement is concerned. Considering the legal importance of details in test arrangements, it must be pointed out that it is vital to involve legal expertise from the country in question when planning a Field Operational Test. The overview given here can, furthermore, not substitute legal advice in a particular case.

A.1 Legal definition of "Field Operational Test"

Before the legal issues related with FOTs are discussed in detail, it must be defined that for the purpose of this section on legal and ethical issues, FOT is considered to be a test arrangement that is accomplished within real-life traffic conditions. This implies that an unknown number of persons not involved in the actual testing procedure, form the surrounding traffic. Usually third parties will not even know about testing being performed. Thus this first characteristic feature excludes artificial, isolated test arrangements and has important legal implications over all legal issues in question.

A.2 Information for test persons (briefing) / contractual agreements

The legal relationship between the organisation carrying out the field test and the test person will most likely have to be agreed upon in a contract.

A further characteristic feature of field operational testing is the data acquisition. In some way the driving will be recorded, possibly even the location might be tracked or videos of the driver and/or surrounding traffic recorded. This has an influence in terms of test persons' and third persons' data privacy and will partially be subject to consent on their side.

A.2.1 Preliminary considerations

To give substantial legal advice on exactly which information must be provided to test persons and estimate which arrangements (insurance, exceptional licenses, etc.) are necessary, rather detailed knowledge on the experimental setup and the systems to be evaluated is required. In order to conclude a contract with test persons, very detailed knowledge on the exact test design and testing procedure is necessary (and this at an early stage of the FOT).

As certain legal consequences may turn out to be unwanted, the following should be taken into account right from the beginning in order to adjust the design of the FOT accordingly.

A.2.2 Information provided to test persons

In order to obtain a valid consent from the driver to log the data and allow for safe participation in the FOT, information on the testing procedure and setup must be comprehensive. This does not include any technical knowledge about the layout itself, but all the consequences for the user: This will make the provision of information necessary on which kind of data is being logged in the first place and who will have access (esp. in case of accidents, administrative fines, etc. this will be important (see sections A.2.3; A.3; A.4.4; A.4.6). Possible legal consequences in case of dangerous driving by the test person should be specified (like possible recourse in case of grossly negligent or intentional behaviour – if applicable (and considered appropriate)). Special attention must be paid towards boundaries and how to deal with malfunctions (the two must be distinguished, see sections A.2.2; A.2.3). The same must be considered in case of a possible overload of the driver in terms of information, warnings, etc. That the responsibility for safe use as for administrative fines remains with the driver must be pointed out explicitly (see section A.3) as must be the fact that the driver remains fully responsible for his/ her driving and is not exempt from full responsibility due to participation in the FOT (this may not fully be applicable in case of non-overrideable systems, however, this will lead to many further questions, see section A.6.2). (Regan, 2006, volume 2)

Apart from this, the effect of the systems on driving – especially in case of some kind of unusual interference – should be pointed out, too. (This might be the case e.g. when applying a visual, acoustic or haptic warning-strategy that might unsettle the driver. (Regan, 2006, volume 2) Special information is particularly advisable in case of any interference into steering or braking, etc.). The risk here is to not sufficiently prepare the test person for safe use of the system. This may under certain conditions lead to liability of the responsible research scientist/ head of department and possibly to liability of the organisation (if possible according to national tort law). Apart from this, a researcher (even negligently) causing damage to the health of a test person may be considered criminally liable for unsafe test design, insufficient instructions and many further substantial breaches of his responsibility with obviously negative connotations for safety.

As a rule of thumb, about the same information and warnings, etc. should therefore be provided that would be necessary in a driver's manual, in case the system is meant to be used under real-life traffic conditions without further surveillance. Thereby, reasonably foreseeable misuse must be taken into account. The information should be provided in a way that the least informed test-person, who is therefore most exposed to a danger, can drive safely. The provision of information and warnings can, however, be achieved otherwise than in a written manual: Personal briefing, presentations on how to handle a system under certain conditions or the training of drivers with the possibility to experience the functioning and ask questions are legally sufficient as well. The possibility to ask questions at any time later during the FOT should be provided for e.g. by means of a telephone hotline (again in order to avoid insufficient briefing which may result in unsafe use by test persons). Furthermore, it might ex post turn out to be difficult to prove that a certain piece of information (that would have been necessary to avoid an accident) has actually been provided to the respective test person. Therefore it seems advisable to incorporate at least

the most important information into the contract with a test person in form of some kind of notice or to refer to another document that has made the information available (see Regan, 2006, volume 2).

A.2.3 Information on system boundaries

A special issue in the context of briefing is system boundaries. System boundaries are those features of a system that are not a defect but still lead to wrong information or an erroneous intervention due to a lack of overall system intelligence (for further details and examples, see Deliverable D6.3).

As system boundaries always occur in certain situations, they are predictable and will not bring about liability issues as long as they have sufficiently been made clear to the test person. A test person who is able to anticipate the system behaviour in the case of all system boundaries will in this respect be considered well-informed.

A.2.4 Information on possible malfunctions

In case the FOT is taken out to evaluate a premature system, possible malfunctions will usually have to be taken into account too. Most important in case of malfunctions will again be to give test persons all the information necessary. That is first of all to provide the information, that a malfunction can occur and instruct thoroughly how to deal with the resulting situations (Regan, 2006, volume 2). If technically feasible, recognisability of malfunctions should be made possible. In most cases it will be sufficient to provide for the possibility to switch off the system and thus ensure safety. Even this might not be necessary as long as the malfunction will not impair safe driving at all. In case of intervening systems, however, much depends on the period of time available for a reaction of the driver: If this is too short, safety will potentially be impaired by any (disturbing) intervention.

For details on possible malfunctions related to the respective systems' intervention or information concept, see Deliverable D6.3.

A.2.5 Information on data recording

As far as data privacy is concerned, details are provided in a separate section (see A.4). For the briefing of test persons it is important to point out the relevant issues for data processing as well as access rights (Regan, 2006, volume 2).

It is legally required that the driver knows which data is being logged. It should also be pointed out, which conclusions can be drawn from the data available and this should involve all imaginable data sources and their combination (including external sources that can be resorted to). The meaning of anonymisation and pseudonymisation as well as any other measures to achieve data privacy should be described too. In case de-personalisation of data is possible and intended, it must be pointed out at which point of data-handling this is realised.

Example a: Within an FOT, data on speed as well as location is recorded. It is possible to anonymise the data for scientific use. However, when logging the data in the car, it can naturally be traced back to the driver (even if personal information is not logged). In this

case, anonymisation might come into effect as soon as the data is read into a database with many similar recordings so that traceability of the test person is barred. Traceability would also be barred, in case the advice in section A.4 is applied. However, the risks of accessibility in spite of these measures (see section A.4.6) – e.g. until pseudonymisation has been realised – should be pointed out too (in order to avoid incomplete information).

The most important measure to comply with data privacy regulations will be to inform the test person thoroughly as far as data acquisition is concerned and (voluntarily!) gain his/ her consent. This consent must – due to the considerable impairment of data privacy combined with FOTs– be stipulated in written form. For all further details and advice, see section A.5 in this handbook (and details in the related Deliverable D6.3).

A.2.6 Agreements on cost allocation and liabilities (including insurance issues)

Another important aspect in terms of contractual agreements is the allocation of costs as well as special agreements in terms of liability. Some aspects will most certainly be regarded appropriate; some might seem disadvantageous in light of volunteer recruitment for an FOT . However, the possibilities in terms of contractual agreements are broad as long as true freedom of decision is ensured (and participation must be voluntary anyway).

Appropriate agreements within an FOT will e.g. presumably be agreements on the allocation of fuel costs that will be borne by the test person. It may furthermore be regarded adequate to agree on a certain sum per mileage for the use of the test vehicle (as long as the vehicle can be employed in everyday use). This again may be combined with other agreements – in case of long term testing – within a lease contract, etc. (Regan, 2006, volume 1).

Of great importance in so far will be the agreements concerning the presumably valuable equipment for data acquisition (and possibly the units installed for evaluation). Here agreements on liability might be necessary as might be a special insurance in order to avoid a financial strain on the test person (and solve this conflict pragmatically). Compare section A.5.

Special agreements will be necessary on data provision by the test persons. As this will mostly be personal data, it shall in so far be referred to section A.4. However, it should be noted that apart from all the agreements necessary in terms of data privacy itself, agreements will also be necessary on how often data shall be retrieved, how this shall take place and the whereabouts of e.g. vehicle return, possibly the demounting of data acquisition components or systems (in case the vehicle remains in the property of the test person). In the latter case special attention must also be paid to possible damages brought about by installation of the FOT-equipment and how these shall be dealt with (Regan, 2006, volume 2).

In case the vehicle does not belong to the property of the test person, special agreements might be necessary in order to assure that the car is not used for dangerous driving. This will be evident from the data retrieved and in severe cases an obligation to intervene might even be brought about as the knowledge on the side of the researcher is evident (and the

participation in an FOT might even provoke dangerous driving depending on the test-persons character). Therefore an appropriate contractual obligation may be stipulated by agreeing on immediate termination of testing, in case dangerous driving is observed (Regan, 2006, volume 2). However, it must be pointed out that this knowledge on dangerous driving usually belongs to the private sphere of the person concerned. In case this knowledge would e.g. be disclosed towards the employer, this would severely compromise the test person and must therefore be dealt with in compliance with the guidelines depicted below (see section A.4). Any disclosure to third parties must therefore – all the more – be refrained from.

Another important aspect in terms of liability of the researcher is to ensure that the test-person is fit to participate in the FOT. This will definitely not imply any detailed inquiries as far as health is concerned (and this would even be considered intimate knowledge in terms of data privacy). Yet, the researcher should not allow a test person to participate in case an unfavourable medical condition is obvious. Apart from this, it might be a good idea to enlighten the need of good health in the FOT -information provided (especially e.g. as far as eyesight is concerned) and this might also be included in the contract dealing with all the details of FOT-participation. The same is true for any substance abuse (see Regan, 2006, volume 1).

Furthermore, information should be provided on the insurances concluded for the test vehicle (in order to point out remaining risks). Depending on the FOT model chosen, this might, of course, only be a recommendation to the test person on which insurances should be concluded (and may even be left completely to the test person in case the systems to be evaluated are mature, in no way critical and the test person owns the vehicle participating in the FOT). Special attention must be paid towards the insurance of data-logging equipment and special agreements might have to be made/ insurance issues pointed out to provide for sufficient information (see section A.5.3).

A.3 Administrative fines

In Germany, administrative fines are related to the personal responsibility of the perpetrator. If traditional driving is considered, no doubts exist on whether responsibility for any breach of traffic law remains with the vehicle driver. However, even systems that only provide information to the driver, tend to point out that traffic rules and traffic signs are prior to information provided (e.g. navigation systems). Therefore the following possibilities for system-design must be considered separately:

A.3.1 Informing Systems

As far as informing systems are concerned, two different types of information must be distinguished: First of all, information may be (more or less) legally irrelevant (e.g. a system providing information on present fuel consumption). Often, especially in case of safety-relevant ADAS, the information will, however, be legally relevant after all. Here again it must be distinguished: On the one hand there is information e.g. on legal speed limits, sign-posted dangerous bends or information provided by road traffic codes. This kind of information has a direct legal implication as it is directly linked to the provisions of road traffic and thus to the conduct legally required. On the other hand, the information that lacks this direct link

may become legally important e.g. in terms of a compensation for damages. The latter is, however, much subject to the contractual agreement and information provided to the test person (for further information on this, see section A.2).

As far as those informing systems are concerned that have been circumscribed to be directly linked to the provisions of road traffic, the question might arise, whether false information provided by the system will excuse or charge (as the case may be) the driver in terms of an administrative fine.

Example A.4.1a: The driver negligently misses a sign-posted speed limit of 50 km per hour at the road side. His car is equipped with a speed alert system so he checks the speed limit displayed there. For some reason the information provided is, however, wrong, a speed limit of 70 km per hour is displayed. The driver relies on the information of his system and drives at 70 km per hour. The driver is fined for speeding.

In terms of administrative fines, it does not matter how the driver has been instructed (at least, if the driver has been aware of the fact that he must generally comply with traffic rules when taking part in the FOT – this information must be provided to the test person, see Section A.2). In example A.4.1a it can be expected from the driver to adhere to traffic signs: Only traffic codes and sign-posted traffic information are legally relevant (no in-vehicle-applications have been introduced in a legally relevant way so far). Because the driver misses the sign-posted speed limit negligently, he can be charged for speeding. All the other information (such as the display of the speed alert system) has no legal implication (it is only a factual “add-on”-information). So even though the driver in example A.4.1a only relies on the wrong information displayed, this will not excuse him legally in a way that the fine cannot be imposed on him (ALBRECHT, 2005).

Example A.4.1b: The driver is speeding and is additionally warned by a speed alert system that he is going to fast. Due to data collection in the car, the display and acoustical signal of the speed-limit warning is recorded. As a camera is also installed, it can be proved that the driver has noticed the warning provided on the display. The driver does, however, not reduce his speed and is fined.

In example A.4.1b the driver is – apart from the sign post or general traffic rule – additionally warned by the in-vehicle-application (such as a “speed alert” system) and has obviously been aware of the speed limit. Therefore his breach of traffic law might be considered intentional, which may have effect on the height/amount of the fine: In Germany e.g. it is generally assumed that speeding is a negligent act. In case intention can be proved – which would be promising given all the data recorded here – the fine will turn out to be higher (ALBRECHT, 2005). This problem is also dealt with in Section A.4 (data privacy issues) as far as data usage in terms of prosecution is concerned. In case it proves to be necessary to record this data, the test person concerned (driver) must at least be aware of the risk he/ she is running (which is again subject to the information provided by the organiser of the FOT).

A.3.2 Intervening, overrideable systems

As far as intervening, overrideable systems are concerned, most important is to point out that they must be actually overrideable in any case and at any time (otherwise see Section A.3.3). If they are overrideable, the driver is still fully responsible for every movement of his vehicle. Usually the intervention will either serve as a basis for information transmission (e.g. vibration of the steering wheel) or will simply intervene by carrying out (a part of the) driving task automatically (e.g. an Adaptive Cruise Control). As far as the transmission of information is concerned, the same will apply in terms of fines (this has been discussed in section A.3.1, see above).

In case the driving task is partially carried out automatically, the system boundaries and functioning of the system must be made completely clear (to provide for full control over the vehicle) and it must be pointed out that the responsibility – even for the aspect of the driving task carried out by the system – remains with the driver. It is therefore necessary to override the system, if this is legally required (and this must, of course, be possible!). All this is subject to the information provided to the test person (see section A.2). As full control over the vehicle will then still be immediately available, administrative fines can be imposed on the driver in case of a negligent or intentional breach of traffic law.

A.3.3 Intervening, non-overrideable systems

In case of intervening, non-overrideable systems, it should for the means of this handbook briefly be pointed out that these are generally considered non-permissible and call for exceptional licenses and a specific insurance (see section A.6.2 and A.5.4).

Apart from this, the driver is no longer capable of (fully) putting his will into execution as far as the control over his vehicle is concerned. In so far as the administrative fine arises from an aspect that no longer belongs to the drivers' control, the breach can no longer be considered negligent or intentional. Therefore administrative fines can no longer be imposed on the driver. (ALBRECHT, 2005)

A.3.4 Cooperative Systems

In case of cooperative systems many aspects may come to effect that have been discussed above in the sub-sections A.3.1, A.3.2 and A.3.3. The respective effect a cooperative system has within the car and the information on the cooperative system provided to the driver will then be respectively valid here. In other words, the same will then apply to cooperative systems.

A.4 Data privacy

A.4.1 Introduction/ general comments/ minimum standard within the EU

Data privacy is in Germany based on basic (constitutional) human rights (for Germany: Art. 1 para. 1 and Art. 2 para. 1 of the "Grundgesetz"= German constitution). This right is termed "informational self-determination" ("informationelle Selbstbestimmung"). The Federal Constitutional Court of Germany characterises this basic right as the authority of the individual to decide on the disclosure and use of his/ her personal data. This is substantiated with the argument that who cannot overlook which personal information is available in

certain fields of his/ her social environment and therefore cannot estimate the knowledge of contact persons, can be substantially hindered to exercise his personal freedom of free decision and planning. This should well circumscribe the scope of protection data privacy acts bring about. [BfD-INFO 1, 2002]

Within Europe, the minimum standard of data privacy ("data protection") is stipulated by the EU Directive 95/46/EG. This directive was issued in 1995 to ensure data privacy of natural persons in the processing of personal data. The directive describes the minimum standard for data protection that must be guaranteed throughout the EU by national law (the directive itself is generally not directly exercisable). In Germany the directive lead to some modifications of national data protection acts such as the "Bundesdatenschutzgesetz" (BDSG). [BfD-INFO 1, 2002], (ROSSNAGEL, 2003)

The extent of protection by data privacy acts in Germany is rather dense. If therefore the data protection principles valid for Germany can be applied to the design of an FOT, it is most likely that this will be sufficient in terms of data protection for other countries of the EU too. However, it must be pointed out that the following statements can only claim definite validity for FOT s[in Germany. In case of doubt, it seems advisable to contact the national data protection officer (if applicable for the respective country) for further advice (the same applies in case of any specific questions). It must also be pointed out that the standard of data protection can turn out to be lower in other countries, should, however, not drop below the minimum standard described in the EU-directive mentioned above. This minimum standard must be complied with, especially when taking out an FOT within an EU research activity. The minimum standard within the EU directive has also been referred to as far as possible.

A.4.2 Legally relevant data and general measures to ensure data privacy

Data privacy regulations are generally based on basic human rights. Therefore the scope of relevant data is restricted to personal data. Personal data are particulars on personal or factual relations of a defined or definable person. In some European Countries (Austria, Luxembourg, Denmark) even legal bodies are covered by data protection rules, however, as far as of interest for FOTs, data privacy of the natural person is in question (see Sec. 3 BDSG, Art. 3 EU Directive 95/46/EG). For further examples see Deliverable D6.3.

Anonymisation and pseudonymisation are measures to assure data privacy.

Anonymisation is the de-personalisation (a modification) of personal data. The data can then not be traced back to the natural person. However, it must be kept in mind, that complete anonymity cannot be achieved, in case the data is so particular that it will apply to only one person. Whether a data set can be considered anonymous, may be dependent on the number of particulars, the available methodological and mathematical instruments as well as the availability of additional information allowing re-personalisation. Therefore, anonymity must be considered a relative term.

In case of pseudonymisation the name or other identification criteria of a person is modified and replaced by a pseudonym (usually a multi-digit number, nick-name or combination of

numbers and letters, the so called "code"). This will considerably complicate the identification of the person behind a data-set. However, in contrast to anonymisation the re-identification remains possible (and is not restricted to chance, mathematical or methodological instruments). With the help of the key that has been separated from the original data set – possibly a list linking the names to the code – re-identification can be achieved. The protection of privacy is much dependent on how well the separation of the key and data-set is ensured. If the key is destroyed, the data would be considered anonymised. (ROSSNAGEL, 2003)

A.4.3 Sub-constitutional law and general principles

For Germany, the basic right of informational self determination has been further developed in sub-constitutional law. Such are the federal law on data privacy ("Bundesdatenschutzgesetz (BDSG)") as well as respective acts in every single federal state. Depending on the background of the organisation taking out the FOT (company (= private) or public authority) different measures are applied in terms of data privacy. For the means of this report, the description of legal framework will focus on the BDSG as this code is generally applicable in case of data privacy for private organisations (companies, etc.) and valid for all federal public bodies (does, however, not apply to the public bodies within the federal states which are large in number: For these the respective act in the respective federal state is applicable. The provisions tend to be very similar, though). [BfD-INFO 1, 2002]

As a rule of thumb, the provisions are, generally speaking, rather strict in case of data acquisition, processing and use by public bodies and more liberal in case of (private) companies. Speaking for Germany, this leads to the situation that only those private companies, institutions, etc. are subject to the federal law on data privacy that collect, process or use data by means of data processing equipment (automated or not). The use of personal data in any other way e.g. by private entities is not subject to the act in the first place. However, data processing equipment will be the rule for field operational testing, so data privacy restrictions will in so far be applicable.

The basic principle of data privacy provisions in Germany is that any form of data acquisition and processing is interdicted, if not subject to explicit authorization within the same act (or some regulation by special law). [BfD-INFO 1, 2002]

Consent of test persons

Based on the provision that data acquisition and processing is generally interdicted, the most important exception from this rule for FOTs is the consent of the person concerned (see Sec. 4 and 4a BDSG, Art. 7 and 10 EU Directive 95/46/EG). For any consent given in terms of data acquisition, processing or use, the person concerned must:

- make this statement in written form (if not certain circumstances make an other form necessary)
- the consequences must be clarified (intended purpose of acquisition/ processing/ use), including the consequences, if consent is not given

- the consent must even be specially highlighted, if the consent to data acquisition/ processing/ use is issued together with other statements (consent to the use, etc. of data concerning health – which might be of interest for FOTs – will call for a special (separate) consent in this respect)

- consent must always be given voluntarily!

[BfD-INFO 1, 2002]

Principle of purpose limitation

Another important principle is purpose limitation (Sec. 14, 28, 29 BDSG; Art. 6 EU Directive 95/46/EG). This means that data may only be processed for the same purpose it has been collected for (or saved, in case prior acquisition is non-applicable).

However, the act comprises a number of exceptions to this rule. As far as relevant for an FOT, acquisition, processing and use of data is not limited to the same purpose in certain cases (see Sec. 14 BDSG valid for public administration and Sec. 28 BDSG for private bodies). [BfD-INFO 1, 2002] For details on these exceptions see Deliverable D6.3.

Data acquisition (extent and limitations)

The general principle for data acquisition is that data must be collected frankly from the person concerned (and not otherwise). As far as data acquisition is taken out by a public body, this is only possible to such an extent as necessary to fulfil the legitimate tasks (Sec. 4, 13, 28, 29 BDSG; Art. 5-7 EU Directive 95/46/EG). The relevant limitations for private entities in case of an FOT are in so far

- specified in the contract on which data acquisition is based

- restricted to explicit consent in case of intimate and very private data (such as racial and ethnical background, political opinion, religious and philosophical belief, health and sex life). Especially data on the health of a test person might be important for the FOT. The acquisition will most likely prove permissible again for the reason of research: Here data acquisition must meet the legal principle of proportionality (and therefore go conform to an evaluation of the higher and therefore more valuable legal right) (see Art. 8 EU Directive 95/46/EG).

Apart from this, data acquisition is also limited by the principle of data economy (Sec. 3a BDSG). This is to say that no more personal data shall be collected and saved than is really necessary to fulfil the purpose in question (i.e. any unnecessary data compilation shall be avoided). [BfD-INFO 1, 2002]

Technical and organisational measures

An important and rather costly aspect of data privacy is the technical and organisational standard that must be applied. Generally speaking, those measures are necessary that will guarantee the compliance with data privacy (see Art. 16, 17 EU Directive 95/46/EG; Sec. 9 BDSG; Sec. 10 BDSG calls for further technical measures but concerns the automatically generated release order and should not be applicable to an FOT). What this rather general

description can imply, has been stipulated in an annexe to the German data privacy act. The effort needed to ensure data privacy in case of automatic processing and usage is dependent on the character: Intimate data is most strictly protected and forms the core area that may not be impaired at all (will seldom be relevant in case of FOTs), personal, private data is strongly protected and data with a relation to other people that is generally known, is the least critical. The character of the data in question indicates the effort to be applied in order to achieve reasonable care.

General measures to ensure data privacy on the technical and organisational level are described in detail in the Deliverable 6.3.

A.4.4 Data privacy in research activities

Research is in itself a basic right of constitutional weight – as is data privacy, see above. Therefore, an appreciation of both values must be carried out for the case in question. Research is facilitated within data protection regulations such as Art. 6, 11 and 13 Directive 95/46/EG and sec. 40 BDSG.

The regulation in sec. 40 BDSG emphasises the principle of purpose limitation to the object of research and explicitly calls for an anonymisation of data at the earliest possible stage. Until anonymisation can be achieved, the characteristics of the person concerned must be saved separate from the particulars on personal or factual relations and must only be brought together in case this is required by the object of research. Any publication of personal data can only be admissible in case the person concerned gives his/ her consent (if relevant for FOTs in the first place).

However (for Germany), no right of professional discretion has been stipulated in the field of research activities (as existing e.g. concerning confidential medical communication of a medical practitioner). This has important implications within criminal law, see below section A.4.6. This may be completely different in other countries of the EU as the Directive 95/46/EG gives sufficient leeway for a deviant regulation. (ROSSNAGEL, 2003)

In order to achieve data privacy for the test persons in spite of these regulations, it has been suggested [ROSSNAGEL, 2003] to deposit data necessary for re-identification (after having pseudonymised the data) with a bearer of secrets (such as lawyers). Such a bearer of secrets can refuse to release data he/ she is entrusted with. It has further been proposed to store the personal data with the (test) person concerned (ROSSNAGEL, 2003). This implies the risk, not to obtain the data, because the test person might finally decide on wanting not to disclose the personal data at all. However, for an FOT this is an option to be considered: As far as the data recorded is stored within the car e.g. by means of a SD-card etc. this would allow the test person to remove the personal data and take care of it by himself/ herself (until it is handed over to the organisation taking out the research activity). The personal data is thus fully placed at the disposal of the test person up to its voluntary release. The test person is, therefore, free to decide upon the further use or existence of the data (and has every right to destroy the data if he/ she pleases). This will avoid conflicts arising from data acquisition and balance the risks implied to a great extent. From a practical point of view it should technically be ensured that the storage medium is easily accessible and removable

(and it should not be too expensive either, because the test-person would supposedly be required to come up for a replacement in case of the storage medium's destruction).

A.4.5 Video recording

Video surveillance might be of great importance for FOTs: Only knowledge from surveillance of the surrounding traffic as well as the driver can determine the impact of a system and/or reveal why a sudden driver action is necessary, etc.

The conflict arises from the high quality of video data that can be achieved by today's technical possibilities in video recording. Additional processing of this data will therefore often reveal the personal identity. Privacy measures and regulations on data privacy in these cases are existent, but they are difficult to handle, see below. The problem with video data is that it is usually impossible to anonymise or pseudonymise images or even videos. This is why this data is especially 'delicate' in terms of data privacy and comprises great dangers. As such must be considered the threat that video-data might be made publicly accessible over the internet (which would be illegal, if the video contains personal data and the test person has not granted his/ her consent prior to this release). Such a video, however, would only then be of great interest, in case the driver or a third party behaves in a way that calls for voyeurism. As the harm to data privacy of the person concerned (and thus his/ her basic right of informational self-determination) can be tremendous, it must be considered reasonable to delete such a sequence entirely as soon as possible (which is as soon as respective knowledge on the existence of the sequence is available). Otherwise, measures must be taken to secure the data adequately (which would be challenging).

From a legal point of view, it must be differentiated between data acquisition outside the car (surrounding traffic) and data acquisition inside the car. This should apply to the technical requirements as well.

Generally speaking, it is advisable, according to the basic principles described above, to do without any video surveillance (in case this proves possible). And pure surveillance (without any recording) is always the less inculpatory measure. Therefore data acquisition by means of video recording should only be employed if indispensable. However, this will usually be the case with FOTs so the following statements will concentrate on necessary video recordings. (ROSSNAGEL, 2003), BfD-INFO 1, 2002]

Video recording of third parties

Video surveillance by means of optical-electronic devices (video surveillance) is only permissible according to the guidelines provided in Sec. 6b BDSG and is also considered critical within the Directive 95/46/EG (see Art. 33 Directive 95/46/EG that contains a revision clause to enhance data privacy in case of video and image surveillance). As far as Sec. 6b BDSG is concerned, it has been criticised that a differentiation of private (video-) data acquisition and video surveillance by public bodies is not made. The scope of sec. 6b BDSG therefore comprises public as well as private bodies and according to its wording it is only applicable to the "surveillance" of publicly accessible places (which would apply to roads). This, however, does not lead to the conclusion that other recordings are irrelevant in terms of data privacy (this is only the case with really private recordings such as those taken within

a family). Any (other) video recordings must therefore meet the provisions stipulated in sec. 6b BDSG, in order to achieve compliance in case of video-data acquisition by non-public institutions. Data acquisition must therefore comply with the following requirements:

Data acquisition must be taken out frankly. A hidden camera will therefore not be regarded permissible as long as consent of the persons concerned has not been obtained.

A legitimate interest for video-data acquisition must exist (see sec. 6 para. 1, No. 3 BDSG). In case of an FOT the concrete legitimate interest will be the same for which video data must be recorded in the first place (and cannot be done without). In case of research (a value of constitutional weight too, see above and Art. 5 para. 3 Grundgesetz = German constitution) the concrete legitimate interest is implicated by the motivation (research).

The video data must be necessary in order to achieve the purpose identified as the legitimate interest. Here it must be considered that the storage of data must also be indispensable, as any storage of video data is considered ultima ratio. This should be overcome in case of an FOT, as long as there is sufficient need of video data.

The video data must be deleted as soon as it is no longer necessary in order to achieve the legitimate purpose of research, see sec. 6b para. 5 BDSG. The same will apply in case of a protection-worthy interest of a third party which is in conflict with further data storage.

(ROSSNAGEL, 2003), [BfD-INFO 1, 2002]

Video recording of the driver

The situation for the driver differs strongly. Basically, however, the same requirements and regulations are valid in this case. The main difference is that the driver will always have to give his/ her consent to the video recording as the permanent recording within the car is strongly invasive. By no means may the recording be taken out with a hidden camera and without informing the driver beforehand. (see ROSSNAGEL, 2003)

Further care must be taken not to record any video data of other passengers, if this can be avoided by technical means (as usually a legitimate interest in so far will not exist, and, as mentioned above, the principle of data economy must be applied). In case the video recording of other passengers is inevitable, it must be ensured that the camera is well on view and thus obvious that data is being recorded. The designated test persons must in this case further be sensitised to inform any passenger of the recording. The same will, of course, apply to any further drivers of the car. It may, however, for a number of reasons be a good idea to restrict the use of the participating vehicles to the actual test person or provide for a "switch off" of all systems and data logging (for safety and data privacy reasons) in case other drivers use the vehicle. This will not only provide for consistent data (which should be necessary from a scientific point of view) but at once ensure that all the contractual agreements actually reach the driver.

A.4.6 Implications of criminal law

If the FOT goes according to plan, criminal law will not be affected. However, in case of accidents, the data collected might be used otherwise. In this case, personal data of the test

person will have been recorded to an extent that is generally not available in this situation in the first place. As the breach of certain traffic rules may be relevant under criminal law aspects, the data will be of interest for means of criminal prosecution as well. This, at least in case the availability of the data is known; however, due to possibly obvious modifications within the vehicles (perhaps even a camera is on view for video-recordings) further inquiries in this respect seem probable. It must therefore be expected that the data might even be subpoenaed on application of a public prosecutor by a judge. For Germany this possibility is given, see sec. 94 seqq. StPO (Strafprozessordnung = Code of criminal procedure). In this context it must be pointed out that these legal effects will be tolerated in Germany and the recording for research-reasons will not privilege the test-person (i.e. it would not be barred to confiscate the data for the reason of criminal prosecution). And data privacy provisions (for Germany) will not bar the use of this data for the reason of criminal prosecution either.

In case the data is already in hold of the organisation doing the research, it would have to be released anyhow (see sec. 95 StPO) – in spite of the fact that this might mean a moral dilemma for the researcher involved in the FOT. These effects, however, may be largely avoided if the procedures suggested for the means of 'data privacy in case of research activities' (see section A.4.4) are taken into effect.

In this context, it must be pointed out that a suspected person always has the right to remain silent in order to avoid self-incrimination (i.e. the accused is not obliged to cooperate actively in the own conviction: "nemo tenetur se ipsum accusare"). It will therefore not be considered a criminal offense in itself, in case the accused would delete or destroy the data recorded. As far as a civil court, however, will decide on compensation for loss suffered, conclusions can be drawn from the fact that the data has been deleted/ destroyed.

A.5 Insurance

A.5.1 Introduction

This section exemplarily goes into the legal situation for Germany as far as road traffic liability and the associated insurance issues are concerned. This should allow for sufficient insight in this aspect to sensitise for possible arrangements and precautions to be taken as far as the insurance of the test-vehicles is concerned.

A.5.2 Road traffic liabilities in Germany

According to national road traffic liability law in Germany, accident victims can potentially claim for compensation from the "keeper" of a vehicle ("Fahrzeughalter", see below), the driver and the vehicle's insurance.

The "keeper" of the vehicle will usually at once be the legal owner; however, this is not invariably true. From a legal point of view, the "keeper" is generally defined as the person that makes use of the vehicle at own expense, i.e. comes up for the costs and has the capitalised use (HENTSCHEL, 2007). The "keeper" will be liable for any damage to the legally protected interest resulting from the operational hazard. The only (basic) requirements for a claim in so far are that the vehicle was in use at the time the damage occurred (this, however, can even be assumed when parking in public space) and that the vehicle's use (i.e.

its operational hazard) has lead to a damage of the legally protected interest. In case a further vehicle plays a part in the emergence of the damage to the legally protected interest, its respective contribution will be considered too. The same applies to contributory negligence of the damaged person. In so far this applies to the causation giving rise to the damage of the legally protected interest. In a further step the remoteness of further damages incurred that can be traced back to the damaging event are considered too.

Unlike the "keeper", the driver will only be liable in case of fault (e.g. any driving mistake, etc. that leads to a damage of the protected interest). Apart from this, the driver is, generally speaking, liable for the same damages to legally protected interests as the "keeper". If the driver's and "keeper's" liability is given, they will both be jointly and separately liable (together with the insurance, see below) for the damage (a term that describes that the damaged person can decide freely which debtor to claim against for the whole damage – which is then settled between the two or more debtors).

Of course, the damage the "keeper" as well as the driver are liable for, is insured via the same compulsory car insurance. By provisions of law the "keeper" is obliged to contract such an insurance in case he wishes to operate his vehicle on public roads. It is regulated that the contract covers the damage on account of the driver as well as the compensation for damages imposed on the "keeper" (and as long as the contractual obligations are adhered to, no recourse will be taken). In Germany, a direct claim of the aggrieved party against the insurance is admissible according to provisions of law. (ALBRECHT, 2005)

A.5.3 Insurances for road traffic in Germany

In so far as material damages are concerned, it is – according to the situation in Germany — important to distinguish between many different types of insurances. First of all, the compulsory road traffic insurance will cover the damage to the property or health of a third party (automobile third party insurance – as stated above, this insurance is compulsory in Germany and therefore widespread and generally referred to as the "car insurance"). It will, however, neither cover the physical damage to the "own" vehicle nor the damage to the health of the driver and other occupants. As far as the physical damage to the car is concerned, a special insurance can be obtained to cover this (comprehensive insurance/ comprehensive coverage insurance including collision). As far as the health of occupants or other passengers is concerned, a special motor passenger personal accident insurance type exists that will come up for damages to passengers. However, it must be kept in mind that the insurance sum for this insurance is usually restricted (and will generally not be sufficient to adequately compensate for serious injuries, special medical care requirements, etc.). Furthermore, the driver might be excluded in this insurance – here a special insurance may prove necessary (driver personal accident insurance). This, however, is again usually limited to certain insurance sums that may not prove to be sufficient for full coverage. Therefore it may appear to be reasonable in some cases – according to the field test design chosen – to obtain some kind of special insurance tailored to the special needs of the specific field trial (e.g. clinical trials insurance).

Most important in all cases will be to disclose the fact towards the insurance that the vehicle is participating in an FOT (which in general should simply be accepted by the insurance). Insurance rates might rise, however, depending on the systems integrated in the vehicle (subject to the FOT, likely in case of premature systems which might involve additional risks). Disclosing this information and possibly incorporating a respective clause in the contract will be a reasonable method to avoid legal uncertainties as far as insurance coverage in case of an accident is concerned.

Automobile Third Party Insurance

As stated above, this insurance is compulsory by law. The minimum insurance sum for this insurance type is e.g. in Germany fixed for motor vehicles at 2.5 Million Euro in case of damages to health (in case of fatal injuries or more than three persons injured: 7.5 Million Euro) and in case of damage to property even limited to 500 000 Euro (see annexe 1 to Sec. 4 of the obligatory insurance law = "Pflichtversicherungsgesetz"). These, however, do not necessarily cover the whole damage in case of serious accidents and even the maximum compensation sums according to the German road traffic act (which is below the minimum insurance sum) can be exceeded in case the claim is based on the law of torts (in this case a maximum compensation sum no longer exists).

It is therefore reasonable to raise the test persons' awareness to these (general) limitations (see section A.2.3 and, if considered necessary, the test vehicle should be insured to better conditions.

Comprehensive insurance/ comprehensive coverage insurance including collision

It must further be decided on the necessity of a comprehensive coverage insurance. This insurance will replace the material damage to a vehicle even in case of self-inflicted accidents (depending on the contract). The insurance will usually exclude intentional damages and may exclude damages resulting from gross negligence. This insurance is not compulsory. If it is renounced, it must beforehand be decided and agreed upon who will come up for these material damages (this insurance would cover) in order to provide for legal certainty. Possibly this has influence on the information that should be provided to the test person, see section A.2.

Motor passenger personal accident insurance

In case of damage to the passengers, compensation can be obtained within the automobile third party insurance of the injuring party. In case of hit and run accidents, absence of insurance coverage of the third party, etc. compensation can be claimed from the personal accident insurance. It will cover all those damages to health involved from the time the passenger gets in the car until he gets out again (HIMMELREICH/HALM, 2006).

This personal motor passenger accident insurance will generally cover costs for medical treatment as well as the costs involved in the event of motor passengers' death.

This insurance shall not be enlarged on within this report. However, selected aspects shall be pointed out: This insurance will not necessarily cover the damage to the driver (which may, however, be the case). And insurance sums vary strongly. They may not be sufficient to

cover severe health injuries or the costs involved in case of disability. Special attention must therefore be paid towards the maximum sums. Further information on this type of insurance should be gathered in the planning of the FOT (if applicable and considered necessary within the test design chosen). (HIMMELREICH/HALM, 2006). Driver Supplementary Insurance

The Driver Supplementary insurance (= "Fahrerzusatzversicherung") is a fairly new insurance type and is largely unknown. The conditions of insurance may differ strongly. Motor passengers are in Germany (since 1st Aug. 2002) covered by the Automobile Third Party Insurance. This is even the case, if it comes to a self-inflicted accident caused solely by the driver of the vehicle they are occupying. In this case, the passengers can claim for compensation against the vehicle's Third Party Insurance. Therefore today only very particular cases depicted above will leave passengers without insurance coverage.

The driver, however, is the only car occupant who may not be able to claim for damages (or only have a partial claim against the third party's insurance). This Driver Supplementary Insurance will cover these damages as far as compensation cannot be obtained otherwise. This insurance type is generally considered reasonable. (HIMMELREICH/HALM, 2006).

Clinical Trials Insurance

As stated just above in this section, the risk of damage to the health of the driver (who is at once the test person) is severe and must be considered beforehand. Of course, the testing in open traffic will also involve many further risks to third parties which must nonetheless be considered. Therefore the justifiable risk will be limited strongly in the first place.

Shall a greater risk nonetheless be taken (and this be considered otherwise permissible), a clinical trials insurance may be necessary to cover the risks involved. As far as (medical) clinical trials are concerned, this insurance type is common. For the purpose of road traffic such an insurance would have to be tailored according to the specific needs of field operational testing.

Test Equipment Insurance

The data-logging equipment and possibly prototype systems may have to be insured too by means of property insurance. This should be kept in mind when planning a Field Operational Test. This electronic equipment will usually not be covered by the comprehensive insurance/ comprehensive coverage insurance including collision (as maximum insurance sums for common electronic equipment in vehicles will assumedly be exceeded by far).

A.5.4 Insurance issues in case of non-overrideable systems

A basic principle of European road traffic is driver's full control. Therefore non-overrideable systems that influence the driving task, comprise unsolved and complex legal questions that cannot be fully assessed at present (see "Communication from the Government of the Federal Republic of Germany to the European Commission of 27 June 2007" page 6, stipulating conclusions of the eSafety Conference in Berlin on 5th/6th June 2007).

As far as insurance issues are concerned, it must therefore be taken into consideration that any type of insurance known today assumes full driver's control. In case the Field

Operational Test is otherwise regarded permissible (see section A.6 and A.7), special arrangements for specific insurance coverage must be taken into consideration.

A.5.5 Insurance issues in case of cooperative systems

Cooperative systems may comprise very specific insurance issues in case the influence on the driving task is strong: If the cooperative aspect therefore involves any kind of vehicle control, it must further be taken into consideration that all present regulations on road traffic are based on the assumption of a vehicles (and driver's) autonomy. In case control of a vehicle is therefore dependent on other vehicles (the same for road side beacons) it must be assessed whether common liabilities (and thus insurance contracts) will sufficiently cover all possible damages in between the linked ("cooperative") vehicles and towards surrounding traffic. If this is not longer the case, it might turn out to be a legally challenging task to tailor the insurance contract to fit the actual "cooperative" situation.

A.6 Vehicle licensing requirements

A.6.1 Licensing requirements for motor vehicles in general

As long as the Field Operational Test is focussed on the evaluation of applications already approved of as optional or standard fitting of the test vehicles, no vehicle licensing requirements will be in question.

The licensing of a vehicle for Europe is generally taken out by means of type approval according to technical rules and regulations in international law. For Europe the so called ECE-Regulations are binding (and their fulfilment will be considered sufficient for road admission throughout Europe). However, gaining a type approval certificate for a new vehicle type is challenging and costly and can hardly be considered an appropriate approach for field testing. Apart from this, it must be taken into consideration that in practice an approved vehicle will serve as a basis for further system integration in an FOT.

In this case the approved vehicle is modified for the purpose of field testing. These modifications may – much depending on the character of the modifications – lead to the cancellation of the vehicle's operating licence. Whether this is the case strongly depends on national licensing requirements. These may still be in existence (as is the StVZO = "Straßenverkehrszulassungsordnung" in Germany). According to the provisions therein, the operating license will expire in case of certain modifications (see Sec. 19 para. 2 StVZO).

This does not apply, in case the vehicle parts integrated have a general approval of their own (which is further specified) or have already been approved of as they are – fitted to the vehicle (and have thus been included in the operational licence of the respective vehicle). In order to make the legal effects of modifications manageable, the German Federal Ministry of Transport has established a catalogue of possible modifications and their impact on the vehicle's operating licence (this catalogue is not legally binding). (HENTSCHEL, 2007) The catalogue will provide a good overview in terms of challenges to be overcome for field testing according to the modifications envisaged. Generally speaking, minor changes such as the safe integration of a display, a separate power supply or data logging equipment will not

lead to the cancellation of the vehicle's operating licence. The modifications must, however, be made transparent.

A.6.2 Special regulation for vehicle manufacturers

In this context an important regulation shall be pointed out that will partly exempt vehicle manufacturers (in hold of the type approval certificate for the respective vehicle) from special licensing requirements (see sec. 19 para. 6 StVZO). In Germany, a vehicle that is used for testing by the manufacturer and registered as such, will not be deprived its operating licence, if further parts are integrated for the purpose of testing. (HENTSCHEL, 2007) This regulation will, however, not permit the modification of vehicles privately owned and registered.

A.6.3 Licensing requirements of "premature" systems/ applications in general

For the purpose of this report a "premature system" shall be considered a system that has so far not been approved of within vehicle type approval and is not separately approved as car accessory either. In order to evaluate such a "premature system" in a Field Operational Test, a special approval might be required to maintain the vehicles operating licence.

For Germany the law within the federal state is decisive as far as the responsibility of the local public authority is concerned (see Sec. 68 StVZO). The responsible public authority will then decide on the necessity of a report by an officially recognised expert certifying consistency with legal provisions. This will usually not be necessary for manufacturers, see above, section A.6.2.

A.7 Special licences (exceptional licences within road traffic law)

A.7.1 Introduction

Further exceptional licences should normally not be necessary for a Field Operational Test – apart from those discussed above in case of modifications on the vehicle's side. This finding will also apply to the drivers' driving licences: Driving licences correspond to certain vehicle types and their use will therefore cover any driver assistance as well as any driver information system that does not put full driver's control into question. And whether data-logging equipment is implemented in a vehicle or not will – if at all – influence the operational licence of the vehicle and does not call for any further special licence.

However, further attention must be raised towards those systems that may intervene beyond full driver's control. In this case, exceptional licences may be necessary after all.

A.7.2 Full control of the human driver

The technology available in the past, as well as the Vienna Convention on Road Traffic (1968) have likewise lead to the assumption of the driver's responsibility to ensure full control over his vehicle. Art. 8 para. 5 and Art. 13 para. 1 of the Vienna Convention explicitly stipulate this full control of the driver over his vehicle "under all circumstances" (Art. 13 para. 1 Vienna Convention on Road Traffic).

The Vienna Convention on Road Traffic formulates a minimum set of requirements in purpose of free (and safe) flow of cross-border transport between the signatory states. The document has had strong influence on the development of national Road Traffic codes and the all-underlying idea of full control of a human driver has thus found its way into many legal provisions concerning road traffic in Germany as well as other countries throughout the EU (and worldwide). In fact, the number of legal provisions based on this idea of full driver's control went without saying and can even be traced back to national road traffic liabilities (which will again influence insurance issues of such systems, see above, section A.5.4).

These findings are common for the EU at large and must be taken into consideration, in case a system shall be evaluated in a Field Operational Test that overrules full control of the driver. In this case, special legal advice as to the consequences the specific system might bring about will be necessary, as will be the application for exceptional licences. These restrictions will, however, not affect systems that do not put the full control of the driver into question. As such must be considered systems that optimise driver initiated functions (e.g. ABS), advisory systems (e.g. speed alert) and fully overrideable ADAS (e.g. adaptive cruise control). Permissible must further be considered those non-overrideable ADAS that have the same effect as traditional technical limits in vehicle performance (e.g. speed limiters) or intervene in situations that cannot be handled by the driver in time and ensure that the intervention keeps in line with the drivers' intentions and will (e.g. ESP and automatic emergency braking). (ALBRECHT, 2005)

Interventions into driving, however, that counteract to the intentions and will of a driver still able to perform the driving task would bring about legal consequences that cannot be predicted at present. (see "Communication from the Government of the Federal Republic of Germany to the European Commission of 27 June 2007" page 6, stipulating conclusions of the eSafety Conference in Berlin on 5th/6th June 2007)

Therefore a need for an exceptional license will arise whenever a non-overrideable system that does not ensure full control of the driver shall be subject to a Field Operational Test.

A.8 Ethical rules

Ethics can be considered as a sub-discipline of philosophy and moral principles guiding behaviour, i.e. they help to distinguish, if a certain conduct is right or wrong. Ethical rules apply and must be obeyed in all kinds of research activities on living organisms and, of course, in particular with human beings. The currently most important ethical rules relevant for research but also professional work on human beings have recently been reviewed as subject of the NoE HUMANIST TF 2 activity on ethical laws and guidelines that apply to behavioural experimental studies (HANZLIKOVA, 2004). However, in the context of the present report and the planning and preparation of FOTsit does not seem to be necessary to review and discuss all principles which apply when e.g. performing medical or genetic research. Here it seems to be sufficient to refer to the key principles for the evaluation of research which are according to (HANZLIKOVA, 2004, p.5):

*"- **Respect for the personality** and his or her autonomy, dignity and self-determination*

Beneficience: *a commitment to maximise potential benefit and minimise possible risks"*

Regarding the planning and performance of FOTs as research projects in the 7th Framework Programme the European Commission makes a clear point when stating: *"All research activities carried out under the seventh Framework Programme must be carried out in compliance with fundamental ethical principles"* (Decision no. 1982/2006/EC; see http://ec.europa.eu/research/science-society). For this reason research proposals shall be evaluated by an independent panel of experts if ethical aspects have been properly addressed. For practical purposes of writing a proposal the EC provides a checklist with critical questions which is designed to help proposers to identify possible relevant ethical issues (see http://ec.europa.eu/research/science-society). With regard to FOTs the following two questions on "Privacy" seem to be most relevant:

Does the proposal involve processing of genetic or personal data?

Does the proposal involve tracking the location or observation of people?

In case of "yes" proposers are required to describe at least the procedures for obtaining informed consent from the persons and the procedures for protecting confidentiality. Moreover, the process of anonymisation or encoding of the data shall be described and it has to be indicated, if the data are used for commercial purposes. These aspects will naturally correspond to legal data privacy provisions, see section A.4.

A.9 Conclusions

From an overall perspective on legal and ethical issues for FOTs, a number of aspects must be taken into account in terms of planning and accomplishing of such testing. This is mainly due to the significant difference between normal driving and FOTs which lies in the evaluation of possibly immature systems (as far as legally permissible), the great extent of data logged and the unique and possibly unprecedented situation test drivers will be confronted with in open traffic.

To sum up, prohibitive difficulties neither from a legal nor from an ethical point of view are in so far to be expected. As long as the advice provided in this report is considered, potential risks – as far as presently foreseeable – can either be settled, avoided or safely handled. It must, however, further be taken into account that advice provided herein is void of knowledge on the concrete system design and thus specific dangers that might be implied in case of particular systems cannot be covered. It can be expected that such concrete difficulties – apart from those indicated as delicate in the report – can be overcome, this, however, will call for further support on legal and ethical issues within the concrete FOT.

Annex A references

[ALBRECHT, 2005]:
Albrecht, F. (2005). Die rechtlichen Rahmenbedingungen bei der Implementierung von Fahrerassistenzsystemen zur Geschwindigkeitsbeeinflussung. Published in Deutsches Autorecht (DAR), 2005, P. 186-198, Munich, Germany.

[BfD-INFO 1, 2002]:
Editor: Der Bundesbeauftragte für den Datenschutz, Postfach 200112, 53131 Bonn (2002). BfD-Info 1 Bundesdatenschutzgesetz – Text und Erläuterungen –. Leck, Germany. http://www.datenschutz.bund.de

[HANZLIKOVA, 2004]:
Hanzlikova, I. (2004). Inventory of ethical laws which apply (at national and EC level) during behavioural experimental studies. NoE HUMANIST, Contract N° 504720, Deliverable 2.1.

[HENTSCHEL, 2007]:
Hentschel, P., König, P., Dauer, P. (2007). Kommentar Straßenverkehrsrecht. Munich, Germany.

[HIMMELREICH/HALM, 2006]:
Editor: Himmelreich, K., Halm, W. Authors: Andreae, M., Becker, A., Bergen, A., Bücken, M., Elvers, R., Engelbrecht, A. Euler, D., Halm, M. Hambloch, R., Hunger, J., Jaeger, L., Kääb, O., Karbach, U. Kreuter-Lange, A., Lehnen, P., Lemor, U. Leser, H., Luckey, J., Mahlberg, L., Müller, M., Nickel, V., Oberpriller, R., Roland, F., Richter, A., Rudolphy, B., Schattenkirchner, S., Schmelcher, M., Steinmeister, M., Schwab, H.-J., Veit, G., Wilms, T., Winkler, T. (2006). Handbuch des Fachanwalts Verkehrsrecht. Cologne, Germany.

[REGAN, 2006, VOL 1]:
Regan, M., Triggs, T., Young, K., Tomasevic, N., Mitsopoulos, E., Stephan, K., and Tingvall, C. (2006). On-Road Evaluation of Intelligent Speed Adaptation, Following Distance Warning and Seatbelt Reminder Systems: Final Results of the Australian TAC SafeCar Project. Volume 1: Report. Monash University Accident Research Centre Report 253. MUARC: Melbourne, Australia.

[REGAN, 2006, VOL 2]:
Regan, M., Triggs, T., Young, K., Tomasevic, N., Mitsopoulos, E., Stephan, K., and Tingvall, C. (2006). On-Road Evaluation of Intelligent Speed Adaptation, Following Distance Warning and Seatbelt Reminder Systems: Final Results of the Australian TAC SafeCar Project. Volume 2: Appendices. Monash University Accident Research Centre Report 253. MUARC: Melbourne, Australia.

[ROSSNAGEL, 2003]:
Editor: Roßnagel A. Authors: Abel, R.-B., Arlt U., Bär W., Bäumler, H., Breinlinger A., Brühann, U., Büllesbach, A., Burkert, H., Dembowski B., Dix, A., Duhr, E., Eiermann H., Ernestus, W., Eul, H., Federrath, H., Garstka, H., Groß, T., Gundermann, L., Hansen, M., Hartig, J., Heibey, H.-W., Heil, H., Herb, A., Hillenbrand-Beck, R., Hoeren, T., Holznagel, B., Königshofen, T., Krader, G., Kranz, H.J., Linnenkohl, K., Miedbrodt, A., Müller, U., Naujok, H., Opaschowski, H. W., Petersdorff, U.v., Pfitzmann, A., Poppenhäger, H., Probst, T., Rasmussen, H., Riegel, R., Rieß, J., Schild, H.-H.; Schirmer, H.D.; Schneider, W., Scholz, P., Sokol, B., Sonntag, M., Tinnefeld, M.-T., Tolzmann, G., Topp, C., Trute, H.-H., Wedde, P., Weichert, T., Werner, M., Wollweber, H. Zezschwitz, F.v. (2003). Handbuch Datenschutzrecht, Die neuen Grundlagen für Wirtschaft und Verwaltung. Munich, Germany.

FESTA Handbook

Version 5

Annex B FOT Implementation Plan (FOTIP)

Revised by FOT-NET

(Field Operational Tests Networking and Methodology Promotion)

28 February 2014

Annex C FOT Implementation Plan (FOTIP)

(To be read in conjunction with Chapter Errore. L'origine riferimento non è stata trovata. of the FESTA Handbook**)**

FOT Teams and People

1. Research Institute contracted to run FOT
2. Project Manager
3. Research Team
4. Technical Support Team
5. Administrative Support Team
6. Project Steering Committee
7. Project Management Team
8. Accounting/Auditing Advisor
9. Legal and Ethical Advisors
10. Sub-Contractors
11. Public Relations and Communications advisor
12. Project Sponsor(s)

Activity 1: ***Convene FOT teams and people***		
Tasks and Sub-Tasks	**Person/ Team/ Organisation Responsible for Activity**	**Done**
1.1 Appoint FOT project manager	Research Institute contracted to run FOT	☐
1.2 Appoint research team	Project Manager,	☐
1.3 Appoint technical support team	Project Manager, Project Steering Committee	☐
1.4 Appoint administrative support team	Project Manager	☐
1.5 Appoint team leaders in each of the research, technical and administrative teams	Project Manager	☐
1.6 Appoint project steering committee	Project Manager, Project Steering Committee, Project Management Team, Public Relations and Communications advisor, Project Sponsor(s)	☐
1.7 Appoint project management team (for –day-to-day management)	Project Manager	☐
1.8 Appoint accounting/auditing advisor	Project Manager, Project Management Team	☐
1.9 Appoint a legal and ethics advisor.	Project Manager, Project Management Team	☐
1.10 Appoint sub-contractors	Project Manager, Project Management Team	☐
1.11 Appoint a public relations/communications advisor	Project Manager, Project Management Team	☐
1.12 Sign off on agreed research and support structure.	Project Manager, Project Management Team, Administrative Support Team Accounting/Auditing Advisor, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts") [1]

- ✓ While the project manager must have knowledge of all activities, ensure that critical knowledge is not vested in just one person. Personnel, including the project manager, may leave the project. *Ensure that there is "standby" for all key research and management roles within the FOT.*
- ✓ Appoint early someone to deal with human participants/ethics committee issues.
- ✓ Include in the research team someone who is a "gizmo" expert – who has up to date knowledge about current ICT/ITS developments and capabilities. Civil engineering and geographical information system (GIS) expertise is also critical.
- ✓ Ensure the project management team meets regularly (about once a month) to resolve research issues, monitor timelines and budgets, and resolve administrative, technical and other issues.
- ✓ *Choose contractors that can guarantee that, if a staff member leaves or is ill, there is sufficient expertise and capacity to maintain project continuity.*
- ✓ Maintain good relations with other partners involved in the FOT.
- ✓ *Ensure that the FOT evaluation process will be, and be recognised as, independent.*
- ✓ It is not necessary to appoint all teams/people at the same time – appointments should coincide with project needs, It is however necessary that the project management team is able to ensure the coverage of the different aspects, eventually asking for support to specific experts in specific domains.
- ✓ Identify a final internal arbiter, acceptable to all parties, who can resolve scientific, administrative, legal and other disputes.
- ✓ Decide early in the project the frequency and timing of project Steering Committee meetings

General Advice [10]

- ✓ Although this Activity precedes Activity 2, the choice of teams and people will be determined to some extent by the aims and objectives of the FOT.
- ✓ Appoint a project manager with excellent research, project management and communication skills. (Note. In some FOTs, the FOT project manager is responsible for both the administrative and scientific management of the FOT. In other FOTs, a senior researcher may be responsible for the scientific, but not the administrative, management of the FOT. This requirement will depend on the scale of the FOT.)
- ✓ The research team should be multi-disciplinary and would typically include psychologists, civil, mechanical, electrical and electronics engineers, statisticians, human factors experts, traffic safety experts, and socio-economic modelling experts.
- ✓ The technical support team would normally include computer software engineers, communications engineers, mechanical, traffic, civil and electronic engineers, and GIS experts.
- ✓ The project Steering Committee sets the strategic direction of the project and keeps it aligned with the project aims and objectives. Normally it would include the FOT project manager, selected members of the research and project management teams (e. g., the team leaders), along with key stakeholders and the sponsor(s). Members should have authority to commit their organizations to the aims, objectives and implementation of the FOT. For smaller FOT projects, the stakeholder committee may not be necessary.

[1] Italic is used to emphasise the most important items

- ✓ The project management team is led by the FOT project manager and includes selected members of the research (e. g, the team leaders), technical and administrative teams.
- ✓ A legal advisor should support the FOT over the full duration of the project (a lawyers office providing advice whenever needed is sufficient). Legal knowledge must be available on the legal situation in the country in which the FOT is conducted.
- ✓ Define all necessary steps including all activities as soon as possible. Even little things can cause big problems
- ✓ Adapt regularly planning of analysis task according to deviations from original project planning and elaborate consequences on analysis task
- ✓ Open communication with regards to project monitoring in order to reduce expectations, if things go wrong in the implementation and data acquisition part
- ✓ It is recommend that experts of phases not immediately operating (e.g. Data Analysis) are involved in the overall FOT/NDS process from the very beginning, not only focusing on the specific task.
- ✓ The same is also recommended involving the right people in the interactions and iterations between planning, using, and analysing
- ✓ The status of project should be diffused to all relevant team members also the ones not yet directly involved.

Activity 2:

Define aims, objectives, research questions and hypotheses

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
2.1 Define aims and objectives of FOT, in conjunction with relevant stakeholders	Project Manager, Research Team, Technical Support Team, Project Steering Committee	☐
2.2 Identify systems and functions to be tested	Project Manager, Research Team, Technical Support Team, Project Steering Committee	☐
2.3 Identify use cases/ situations in which systems and functions are to be tested	Project Manager, Research Team, Technical Support Team Project Steering Committee	☐
2.4 Define research questions and prioritise them	Project Manager, Research Team, Technical Support Team Project Steering Committee	☐
2.5 Formulate hypotheses to be tested, deriving from research questions	Project Manager, Research Team, Technical Support Team Project Steering Committee	☐

2.6 Determine constraints which may prevent the aims and objectives from being met	Project Manager, Research Team, Technical Support Team Project Steering Committee Project Management Team	☐
2.7 Define final aims and objectives of the FOT, and seek agreement from relevant stakeholders.	Project Manager, Research Team, Technical Support Team, Project Steering Committee	☐
2.8 Sign off on aims and objectives of FOT	Project Manager, Research Team, Technical Support Team, Project Steering Committee, Project Management Team, Public Relations and Communications advisor, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ Be prepared for the potential for FOT aims and objectives to change when new administrations come in.
- ✓ Be prepared for the potential for conflict in objectives by different stakeholders. e. g., a car manufacturer wants a deep understanding of product use and driver behaviour and acceptance while public authorities are more interested in determining the impact of system use on traffic and on the transport system.
- ✓ Check that use cases may be practically tested and reproduced with the requested statistical relevance in order to avoid that unfeasible use cases (and related research questions and hypotheses) may be discarded in the successive steps.
- ✓ Ensure that all terms and phrases making up the research questions and hypotheses are clearly defined and unambiguous. This will facilitate interpretation of the FOT outcomes and comparisons with previous and future FOTS.

General Advice

- ✓ See the FESTA Handbook for further advice on defining the aims, objectives, research questions and hypotheses for an FOT.
- ✓ Constraints which may prevent the aims and objectives from being met might include cost, lack of supporting infrastructure, time, willingness and commitment of key stakeholders to cooperate in providing supporting infrastructure, their likely support in promoting the aims and objectives of the FOT, the availability of appropriate data, etc
- ✓ Commonly cited aims are:
  - evaluate system(s) effectiveness in changing behaviour and performance
  - evaluate driver acceptance of system(s), including willingness to purchase
  - evaluate system technical operation
  - stimulate societal demand for new technologies
  - evaluate safety impacts
  - evaluate environmental impacts
  - evaluate impacts on traffic (e. g., congestion, mobility)
  - evaluate socio-economic cost-benefits

- evaluate commercial impacts (e. g., productivity, return on investment, direct cost savings, incremental revenues by getting more customers, customer loyalty, etc.)

✓ *Defining the research questions and prioritizing them at an early stage will ensure they stay at the focus of the FOT and help protect from subsequent "mission creep"*

Activity 3:

Develop FOT project management plan

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
3.1 Define project activities, tasks and sub-tasks	Project Manager, Project Management Team	☐
3.2 Decide who is accountable for completion of activities, tasks and sub-tasks	Project Manager, Project Management Team	☐
3.3 Determine timelines for completion of activities, tasks and sub-tasks	Project Manager, Project Management Team	☐
3.4 Determine budget for project activities, tasks and timelines	Project Manager, Project Management Team	☐
3.5 Develop a project GANTT chart to guide project management	Project Manager, Project Management Team	☐
3.6 Implement procedures for monitoring project activities, timelines, budgets and resources (e. g., project management team meetings)	Project Manager, Project Management Team	☐
3.7 Undertake a risk assessment for the FOT and plan contingencies as required.	Project Manager, Project Management Team, Risk Management Consultant	☐
3.8 Determine sign off procedures (meetings and documents) to ensure that there is sign off on all critical decisions and stages in the FOT by all relevant parties	Project Manager, Project Management Team	☐
3.9 Agree on project issues which are confidential and implement mechanisms for safeguarding their confidentiality.	Project Manager, Project Management Team, Project Sponsor(s)	☐
3.10 Develop a manual for conducting the FOT that documents critical procedural knowledge.	Project Manager, Research Team, Technical Support Team	☐
3.11 Sign off on project management plan.	Project Manager, Project Management Team, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ *Include in the total budget some "contingency" that can be used to pay for unforeseen activities and tasks (especially meetings) that cannot be anticipated.* 5 -10 percent of the total project cost is recommended. Different elements of the project may require different proportions of this contingency. It should be held and allocated by the project manager, not sub-activity leaders or partners.
- ✓ Identify and document in the GANTT chart the dependencies that exist between different activities, tasks and sub-tasks.
- ✓ Anticipate the need and budget for specialist consultants with skills and expertise that does not exist within the project team (e. g., training experts, software developers, lawyers etc)
- ✓ Where relevant, anticipate changes to 3rd party vehicle fleets (e. g. vehicle upgrades and changes in operating routes) during the course of the FOT.
- ✓ Be aware that technical efforts are most likely to incur risk in terms of time and budget (especially the hardening up/refinement of systems, where these are developed within the FOT)
- ✓ *Don't under-estimate the time required and the cost of designing, running, analysing and de-commissioning the FOT. It will be greater than you think.*
- ✓ Assume that some further modifications to, and fine tuning of, the project management plan will be required. It is impossible to foresee everything that is required in running an FOT.
- ✓ Develop procedural manuals for those conducting the FOT to ensure that, if staffs leave, all procedural knowledge does not leave with them. These should be developed for each activity.

General Advice

- ✓ *Documentation of all project meetings is critical to record critical decisions, document the lessons learnt and justify possible blowouts in budgets and timelines.*
- ✓ A budgeting structure that accommodates the uncertainties associated with running FOTs is desirable – for example, a series of prospective budgets for each critical stage of the FOT.
- ✓ *Be aware that in some jurisdictions project papers from publicly funded projects are public documents and copies can be requested by members of the public.*

Activity 4:

Implement procedures and protocols for communicating with stakeholders

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
4.1 Commission communications advisor to design communications plan	Project Manager, Project Management Team	☐
4.2 Develop and implement communications plan	Project Manager, Project Management Team, Public Relations and Communications advisor	☐
4.3 Appoint media spokes people	Project Manager, Project Management Team, Project Steering Committee	☐
4.4 Sign off on agreed communication protocols.	Project Manager, Project Management Team, Public Relations and Communications advisor, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts") ✓ *Assume that you will be mis-represented by the media. Try and limit media attention until the data collection is complete.* ✓ *Agree in the contract with the sponsor who is responsible for press releases and dissemination of information and results.* ✓ *FOTs attract a lot of media attention. Provide adequate time and budget for unsolicited communication with stakeholders, especially with the media.* ✓ Ensure that the project steering committee has input to the communications plan. ✓ *Ensure that there is appropriate control of communication with the media, through the appointed media spokesperson. For EU projects, involving multiple partners, it may be necessary to appoint more than one media spokesperson.* ✓ Everyone involved in the project must know who the media spokesperson is. ✓ *The media spokesperson should consult with the project management group before speaking to the media, especially on sensitive issues.* ✓ Provide media training for appointed spokespeople. ✓ *Build political support for the FOT early in the project, and maintain it during and after the FOT.* ✓ *Be aware that there may be some key stakeholders who believe that FOTs are an impediment to system rollout. These people, in particular, must be made aware of the rationale for FOTs.* ✓ Plan to have some results available at early stages of the project. If desirable, they should be released to an informed audience (e. g., at a conference) but not to the media as they could contaminate subsequent data collection ✓ Plan for annual public meetings, and a project website, to disseminate information and findings. ✓ *Don't undermine the scientific integrity of the research program by mis-timing communications with the media and other stakeholders* ✓ *Have a response prepared in case of serious incidents – such as a crash involving a test vehicle. Anticipate media contact between the media and participant drivers.* ✓ Be aware that fleet/truck drivers may be more inclined to disclose opinions to the media if asked.		
General Advice ✓ *Open communication with key stakeholders is important at an early stage of the FOT to ensure that the aims and objectives of the FOT are clear, that stakeholders are committed to the project, and that the aims and objectives of the FOT are not misquoted, misrepresented or misunderstood.* ✓ There should be an agreed minimum level of transparency and result sharing in the FOT — avoid "confidential FOTs". ✓ It may be beneficial to engage a professional press office to handle external communications, particularly with the media. ✓ FOT drivers and FOT researchers are usually of most interest to the media. ✓ Decide in advance with stakeholders a minimum time for approval for statements released to the media. ✓ Be prepared for the possibility that politicians may at times want to veto communications between the FOT project team, the media and other stakeholders. ✓ Building political support outside the project can help provide protection against strong partners/sponsors ✓ *Early negative media attention may have a significant impact on participant recruitment and/or colour participant expectations of system performance. Try to prevent any media*		

awareness until after the recruitment phase is complete.		
Activity 5: ***Design the Study***		
Tasks and Sub-Tasks	**Person/ Team/ Organisation Responsible for Activity**	**Done**
5.1 Become familiar with the methods, measures and procedures of previous FOTs: **o Read the FESTA handbook** **o Attend the FOTNET seminar and similar events and networking activities** **o Talk to experts who have conducted FOTs previously.** **o Review the relevant literature**	Project Manager, Project Manager, Project Manager, Technical Support Team	☐
5.2 Identify the performance indicators necessary to test the hypotheses derived in Activity 1	Project Manager, Research Team, Technical Support Team	☐
5.3 Select measures (objective and subjective) that allow performance indicators to be derived to test the hypotheses	Project Manager, Research Team, Technical Support Team	☐
5.4 Identify the sensors and sensor requirements for obtaining the required measures	Project Manager, Research Team, Technical Support Team	☐
5.5 Design the experimental methods, tools and procedures for testing the hypotheses	Project Manager, Research Team, Technical Support Team	☐
5.6 Define methods, tools, requirements and procedures for acquiring, storing, transferring, de-coding, reducing/transcribing, filtering, backing up and verifying the data	Project Manager, Research Team, Technical Support Team	☐
5.7 Define methods, tools and procedures for analyzing the data	Project Manager, Research Team, Technical Support Team	☐
5.8 Determine optimal sample size (conduct power analyses) to ensure sufficient statistical power.	Project Manager, Research Team	☐
5.9 Select models for estimating the potential safety, environmental and other benefits of the technologies tested.	Project Manager, Research Team	☐
5.10 Sign off on study design, methods and tools, questionnaires and associated procedures.	Project Manager, Project Management Team, Project Steering Committee, Legal and Ethical Advisors, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts") ✓ *Ensure that necessary historic data (e. g., data on vehicle speeds on certain roads) is available for baseline comparisons or Cost Benefit Analysis.* ✓ Where relevant, allow sufficient time between vehicle allocations for system maintenance and verification, servicing and repairs to be undertaken. ✓ Accept that it is impossible to design a perfect FOT. Many practical issues – including time and money — will constrain the final experimental design. ✓ Remember that an FOT is not an experiment – control is limited, and counterbalancing may not be possible. ✓ *Design into the FOT a contingency plan, in case there is an unexpected requirement to reduce or increase the scope of the study (e. g, to save money or time).* ✓ Employ a multidisciplinary team in developing hypotheses that includes researchers and people with expert knowledge about the systems to be tested. ✓ *Design the study in a way that allows for direct comparisons to be made between objective data (logged by the platform) and participative data (collected through questionnaires, focus groups etc).* ✓ Keep to an acceptable minimum the number and size of questionnaires that must be completed by participants at different points of the study, to maximise the likelihood of them being completed. A sub 2-hour completion duration is a useful target, as longer sessions may tend to remind participants that they are part of a scientific study. ✓ *Don't be tempted to reduce the sample size in order to save money – conducting a study with too few participants leads to a lack of statistical power to detect effects, and may ultimately be a waste of time and money.* ✓ Make sure that everyone understands the FOT study design, so that they appreciate the timing issues and the consequences of wanting to make changes to it e. g., if wanting to reduce the scope of the study. ✓ Delays in one area of the program cannot necessarily be made up by making sacrifices to other areas. ✓ *Don't assume that FOT users are the only ones who will use the FOT platforms.* ✓ Don't be pressured into changing the design of the study if, in doing so, it compromises the scientific integrity of the study. ✓ *When performing the sample size calculations, allow for participant attrition, e. g. if using fleet drivers, some may leave the company during the FOT period.* ✓ Where hypotheses are not supported, consider conducting a process evaluation. This can help determine whether the system did not work, or whether any implementation issues may have impacted on the results. Plan for annual public meetings, and a project website, to disseminate information and findings.		
General Advice ✓ See FESTA Handbook for detailed advice on designing the research study. ✓ See the FESTA Handbook reference list for published reports on previous FOTs. ✓ Where it is not possible, for ethical, practical or safety reasons, to investigate an issue in an FOT, consider safe alternative means for doing the research (e. g., simulators, test tracks). ✓ *The level of driver familiarity with the test vehicle may influence driver performance during the early stages of the FOT.* ✓ Ethical incentives that can be given to discourage driver attrition from the study should be agreed on early in the project. ✓ The models for estimating safety and other benefits may need to be updated in response to recent literature when making the estimation. ✓ *For the business sector, the commercial impact of the technologies deployed (e. g., in terms of productivity, return on investment, cost savings, incremental revenues by getting more customers, customer loyalty, etc) will be important to evaluate.*		
Activity 6:		

Identify and resolve FOT legal and ethical issues		
Tasks and Sub-Tasks	**Person/ Team/ Organisation Responsible for Activity**	**Done**
6.1 Seek specialist advice to identify relevant legal and ethical issues	Project Manager, Accounting/Auditing Advisor, Legal and Ethical Advisors	☐
6.2 Resolve all legal and ethical issues that can be identified in advance	Project Manager, Project Management Team, Accounting/Auditing Advisor, Legal and Ethical Advisors	☐
6.3 Create contracts and/or agreements with all relevant parties (e. g., vehicle leasing organisations, suppliers, road operators, traffic centres, consultants, fleet managers, researchers etc) for all relevant issues (e. g., data collection, provision and usage, theft, insurance, privacy, duty of care, property, disposal of vehicles after the study, etc)	Project Manager, Project Management Team, Accounting/Auditing Advisor, Legal and Ethical Advisors	☐
6.4 Seek ethics approval to conduct study (where required) from relevant ethics committee	Project Manager, Research Team, Technical Support Team, Legal and Ethical Advisors	☐
6.5 Seek expert advice regarding liability issues and to ensure insurance provision is adequate for all foreseeable eventualities	Project Manager, Accounting/Auditing Advisor, Legal and Ethical Advisors	☐
6.6 Ensure that vehicle type approval and warranty requirements are adhered to in spite of the modifications (implementation of data logging equipment and possibly systems to be evaluated, etc.)	Project Manager, Research Team, Legal and Ethical Advisors	☐
6.7 Obtain informed consent of participants before they are allowed to participate in the FOT	Project Manager, Research Team, Technical Support Team	☐
6.8 Sign off on all aspects of the FOT design and procedures pertaining to legal and ethical matters.	Project Manager, Project Management Team, Accounting/Auditing Advisor, Legal and Ethical Advisors, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ In terms of the project timeline, legal and ethical issues need to be considered in parallel from the beginning to the end (and indeed afterwards in terms of data protection)., especially if the data are planned to be re-used in new projects.
- ✓ *There must be mutual agreement on the relative risks to all parties before contracts are signed*.
- ✓ *Double check that the final design and conduct of the FOT accords with ethical and legal requirements in all jurisdictions in which the FOT will physically occur.*
- ✓ Ensure that all intellectual property issues are identified and resolved "up front".
- ✓ Ensure permission to drive (and necessary insurance cover) restrictions are understood by all parties, particularly participants.

- ✓ Identify the conditions under which a participant will be expelled from the study, and ensure these are made known to participants before the FOT commences.
- ✓ Ensure that all participating drivers are fully licensed to drive the test vehicles.
- ✓ *Don't forget about the need to adhere to contractual obligations and confidentiality agreements. FOTs often extend over long periods, making it easy to lose sight of obligations and agreements*.
- ✓ Clarify participant responsibilities and the study's obligations to the participants. Participant responsibilities should include routine vehicle maintenance activities e. g., checking fluid levels.
- ✓ *Ensure all relevant health and safety requirements of participants and the study team are met.*
- ✓ for anonymised data to be passed to 3rd parties. (NB with GPS and video data it may be very difficult to guarantee anonymity). All project staff must understand who has access to project data, especially video data.
- ✓ *All study team members must understand the agreed response should a major incident, such as an accident, occurs. Media comment should only be made by the spokesperson.*
- ✓ Don't underestimate the complexity and the time commitment involved in identifying and resolving the legal and ethical issues associated with the conduct of an FOT.
- ✓ *Ensure that all methods, tools, procedures and materials used in the study that require legal and ethics approval are approved by the ethics committee at appropriate points in the study.*

General Advice

- ✓ See Deliverable D 6.3, Annex A and Chapter **Errore. L'origine riferimento non è stata trovata.** of this FESTA Handbook for detailed advice on legal and ethical issues.

Activity 7:

Select and obtain FOT test platforms (vehicles, mobile devices, road side units, ...)

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
7.1 Specify functional requirements, performance specifications and user requirements for the test platforms needed for the study.	Project Manager, Research Team, Technical Support Team	☐
7.2 Specify functional requirements and performance specifications for the integration into platforms of all systems needed for the FOT (FOT technologies, support technologies and data collection technologies), if these are not already in the platforms.	Project Manager, Research Team, Technical Support Team	☐
7.3 Select test platforms (makes and models) that meet above requirements.	Project Manager, Technical Support Team	☐
7.4 Where relevant purchase, lease, hire or borrow (where the driver owns the	Project Manager, Accounting/Auditing Advisor	☐

vehicle) the test vehicles and/or platforms.		
7.5 Sign off on selection and obtaining of test platforms.	Project Manager, Technical Support Team, Project Management Team, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ The choice of platforms may well impinge on the selection of participants which, in itself, will impact on the research questions. Choice of platforms must be undertaken at an early stage in the project's planning.
- ✓ *Consider obtaining extra test platforms. These can be used as spare items in case of failure and as "showcasing" platforms. The latter can be driven at appropriate times by politicians and other high ranking officials in positions of authority to promote and deploy the systems on a wider scale.*
- ✓ Be aware that vehicle choice may affect participant response if the test vehicle is significantly better/worse than the vehicle they are used to driving. Choose a conservative model.
- ✓ *Do consider vehicle maintenance requirements and the dealer network that is available in the FOT area. If the FOT will take place in a limited area, consider advising the local dealer(s) of the study. This may be important if a participant takes a test vehicle to a dealer to fix a problem.*

General Advice

- ✓ Where used, the test vehicle will vary, depending on the nature of the FOT. In some FOTs, the test vehicles will already contain mature OEM systems. In others, the systems will need to be developed (fully or partly) and integrated into the vehicles. In some FOTs, the systems will be integrated into drivers' own vehicles; in others, they will be integrated into company fleet vehicles.
- ✓ The test platforms must be capable of hosting the technologies to be evaluated (OEM, aftermarket and nomadic) and the data logging and support systems
- ✓ Carefully evaluate the trade-off between owned and leased equipments and vehicles to be used in FOTs in order to choose the most convenient solution for the specific responsible organization.

Activity 8:

Select and obtain systems and functions to be evaluated during FOT (if they are not already implemented in the test platforms)

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
8.1 Develop selection criteria for choosing systems and functions (OEM, aftermarket and nomadic) to be tested (if the technologies to be tested have not already been selected by the sponsor; see General Advice column).	Project Manager, Research Team, Technical Support Team, Project Steering Committee	☐
8.2 Use above selection criteria to select and obtain systems to be tested	Project Manager, Research Team, Technical Support Team	☐
8.3 If commercial systems do not exist, that meet the above criteria, develop functional requirements and performance specifications for systems that do, (including	Project Manager, Research Team, Technical Support Team and (if appropriate) consultant	☐

for HMI and security issues).		
8.4 Develop functional requirements and performance specifications for the infrastructure needed to support the deployment of the technologies to be tested (e. g. digital maps, roadside units).	Project Manager, Research Team, Technical Support Team, Project Steering Committee, and (if appropriate) consultant	☐
8.5 Source infrastructure that meets the above functional requirements and specifications.	Project Manager, Technical Support Team and (if appropriate) consultant	☐
8.6 Where infrastructure is not commercially available, develop supporting infrastructure that meets the above functional requirements and performance specifications.	Project Manager, Research Team, Technical Support Team and (if appropriate) consultant	☐
8.7 If appropriate, issue Expressions of Interest/Requests for Tenders for provision of systems and supporting infrastructure.	Project Manager, Project Management Team, Accounting/Auditing Advisor, Legal and Ethical Advisors	☐
8.8 If appropriate select preferred tenderers, negotiate contracts and award contracts.	Project Manager, Project Management Team, Accounting/Auditing Advisor	☐
8.9 Decide what will be done with the test platforms, and the equipment in them, once the FOT has been completed.	Project Manager, Research Team, Technical Support Team, Administrative Support Team, Project Steering Committee, Legal and Ethical Advisors, Project Sponsor(s)	☐
8.10 Sign off on selection and obtaining of systems and functions to be evaluated during the FOT	Project Manager, Technical Support Team, Project Management Team, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ *Do ensure that criteria for the selection of candidate systems (where this is appropriate) to be evaluated are developed in consultation with relevant stakeholders, to ensure the systems to be tested meet the needs of all relevant stakeholders and are suitable for in-car use (this includes good interface design).*
- ✓ Selection of systems must be undertaken with consideration of the data-logging system. If not, problems of interfacing may result.
- ✓ Beware of hidden costs of hardware and software development if these items are not originally designed for research purposes.

General Advice

- ✓ Criteria for selection of candidate systems in the FOT (if they have not been pre-selected by the sponsor) could include: likely safety or environmental benefit, likely benefit in increasing commercial productivity and efficiency, availability, compatibility with host vehicles, technical performance, cost, reliability, maintainability, likely acceptability to drivers, usability, compliance with relevant human factors/ergonomic guidelines, compliance with local legal requirements, compliance with relevant standards, crashworthiness etc.
- ✓ If prototype systems are tested, then estimates of durability, reliability, maintenance costs etc. of production systems will be difficult, and full Cost Benefit Analyses may not be

possible.

Activity 9:

Select and obtain data collection and transfer systems

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
9.1 Specify data to be logged (measures and sampling rate)	Project Manager, Research Team, Technical Support Team	☐
9.2 Specify functional requirements and performance specifications for systems for collecting and transferring the data to be logged.	Project Manager, Research Team, Technical Support Team	☐
9.3 Source, purchase and/or develop systems for logging and transferring the data that meet the above functional requirements and performance specifications.	Project Manager, Technical Support Team and (if appropriate) sub-contractors.	☐
9.4 Sign off on selection and obtaining of data collection and transfer system.	Project Manager, Project Management Team, Legal and Ethical Advisors, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ Implement re-calibration procedures that will ensure accuracy of measurements/sensors over time and help prevent data drift issues.
- ✓ Plan for software upgrade and revision during the FOT and try to ensure that all software systems are updated together. Ideally, this should be possible remotely.
- ✓ *Where used, in-vehicle data logging systems need to be unobtrusive, safe and secure — but they also need to be accessible to enable routine repairs.*
- ✓ *Where relevant, provide a local location for vehicle support and a vehicle tracking capability.*
- ✓ Minimise user involvement in data download from test platforms.
- ✓ *Ensure boot-up time for test systems and data logging systems is sufficiently fast to prevent data loss at the beginning of each trip.*
- ✓ *Ensure that a common time stamp is used for all recorded data sources.*
- ✓ Verify the definition of signals provided by 3rd parties (e.g. CAN message definitions by vehicle manufacturers)
- ✓ Do not allow data collection to proceed automatically without active confirmation of data capture and validity. This may include the generation of warning messages (SMS?) when out of tolerance data is recorded.
- ✓ Recognise that some data is much more important than others and should be given a relatively higher priority.
- ✓ *Do keep a stock of spares for critical items and anticipate that some components may become unobtainable during the study.*
- ✓ Consider the opportunities for ad-hoc and post-hoc interrogation of raw data files to answer additional questions. This may not be possible if data collection is triggered.

General Advice

- ✓ See chapters **Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.** of the FESTA Handbook for more detail.

- ✓ The technologies fitted to test vehicles may also include supplementary technologies (such as sensor technologies; e. g., forward looking radars, GPS) that are needed to, for example, measure inter-vehicle following distances in order to determine whether speeds are free or constrained (e. g., see Regan et al, 2006, Volume 1).
- ✓ See Deliverable D6.3, Annex A and Chapter **Errore. L'origine riferimento non è stata trovata.** on legal issues of data privacy to be aware of possible dangers and legal provisions.

Activity 10:

Select and obtain support systems for FOT platforms

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
10.1 Define the support systems needed (see General Advice Column)	Project Manager, Research Team, Technical Support Team and (if appropriate) consultant	☐
10.2 Develop functional requirements and performance specifications for systems needed to support the study	Project Manager, Research Team, Technical Support Team and (if appropriate) consultant	☐
10.3 Where appropriate, develop functional requirements and performance specifications for the HMI, to ensure that the HMI for support systems is safe and user-friendly	Project Manager, Research Team, Technical Support Team and (if appropriate) consultant	☐
10.4 Source, purchase and/or develop support systems that meet above functional requirements and performance specifications	Project Manager, Technical Support Team and (if appropriate) sub-contractors.	☐
10.5 Sign off on selection and obtaining of support systems for test platforms	Project Manager, Project Management Team, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ If possible, support systems should be capable of remote operation to allow, for example, remote system reboot.
- ✓ In the case of very large naturalistic studies it may not be practicable to intervene manually. In these cases do attempt to automate as much as possible.
- ✓ Anticipate data analysis requirements before specifying data to be logged (e. g., rates and resolution)
- ✓ Ensure that missing data are clearly indicated – e. g., if the data collection system malfunctions, missing data should NOT be indicated with a zero, where zero is a valid measure (e. g., speed).
- ✓ If in doubt about the final list of measures to be logged, log more parameters if performance of the data logging system or storage capacity are not affected. Consider the opportunities for ad-hoc and post-hoc interrogation of raw data files to answer additional questions. This may not be possible if data collection is triggered.

General Advice

- ✓ Support systems have multiple purposes: e. g., to display information to users; to automatically turn systems on and off where multiple systems are being tested and exposure to each is kept constant across drivers; for manually disabling systems in the event of malfunctions (i.e., "panic buttons"); for preventing use of systems by non-participants; for diagnosing system status and faults; etc.

Activity 11:

Equip FOT test platforms with all systems

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
11.1 Prepare a system installation/integration manual describing standardized procedures	Project Manager, Technical Support Team and (if appropriate) sub-contractors.	☐
11.2 Equip test platforms with the FOT systems to be evaluated (if not already installed)	Project Manager, Technical Support Team and (if appropriate) sub-contractors.	☐
11.3 Equip test platforms with data collection and transfer systems	Project Manager, Technical Support Team and (if appropriate) sub-contractors.	☐
11.4 Equip platforms with FOT support systems (e. g., panic button, for turning systems off in a vehicle etc.)	Project Manager, Technical Support Team and (if appropriate) sub-contractors.	☐
11.5 Sign off on system integration activities, ensuring that all systems have been installed in accordance with the system installation/integration manual	Project Manager, Project Management Team, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ Ensure that the computers running all systems (FOT, data collection and support) have sufficient computing power to avoid processing delays.
- ✓ *Ensure that all systems (FOT, data collection and support) operate identically across test platforms.*
- ✓ *Allow all new vehicles a burn-in period (around 1000km) so that vehicle faults, that could disrupt the FOT, can be detected.*
- ✓ Be aware that 'identical' platforms and sensors may perform differently due to variation in components, manufacturing variability and environmental conditions. Check for differences that may be critical for the FOT.
- ✓ Try and make all adaptations to test vehicles (e. g., fitment of novel display systems) invisible to reduce the likelihood of theft or behaviour modification by other drivers.

General Advice

- ✓

Activity 12:

Design and implement user feedback and reporting systems

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
12.1 Design, develop and implement systems and procedures to allow users to report technical problems in a timely manner.	Project Manager, Research Team, Technical Support Team	☐
12.2 Design, develop and implement systems and procedures to allow users to provide feedback to researchers, in real time or retrospectively (e. g., usability problems, opinions of systems, confirmation that systems are operating as required etc)	Project Manager, Research Team, Technical Support Team	☐
12.3 Design, develop and implement systems and procedures that allow researchers to monitor participant progress (e. g., to ensure they are adhering to study requirements).	Project Manager, Research Team, Technical Support Team	☐
12.4 Sign off on implementation of user feedback and reporting systems and procedures	Project Manager, Research Team, Technical Support Team, Project Management Team, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ Implement 'user diaries' to allow confirmation of user identity and trip details if this process cannot be automated. This may encourage users to behave less naturally.
- ✓ Implement a timetable for the timely collection of qualitative data so that participants don't have to rely on their memories.
- ✓ Implement a timetable for periodic verification of collected log data and checking that systems/cars still operate correctly.
- ✓ *Anticipate that users may not complete diaries accurately or consistently and may fail to attend for de-briefing interviews. Appoint user liaison staff as a single point of contact.*
- ✓ *Ensure that the project team can respond to emergencies and incidents on a 24/7 basis.*
- ✓ Do ask participants to announce when they are going on holiday or not using the platform for an extended period.
- ✓ Keep a record of all reported problems, and document these in relevant reports.
- ✓ Ensure that all feedback and reporting procedures are documented in a manual for quick reference by the research and technical support team as required.
- ✓ Consider whether you need to design, develop and implement a system to allow for the collection of fuel consumption information
- ✓ Where fuel consumption is calculated manually, anticipate that drivers will not always use fuel cards, return fuel dockets or fill in the fuel logbook.

General Advice

- ✓

Activity 13: **Select, obtain and implement standard relational database for storing FOT data**		
Tasks and Sub-Tasks	**Person/ Team/ Organisation Responsible for Activity**	**Done**
13.1 Design, develop and implement a database for storing data logged from the test platforms	Project Manager, Research Team, Technical Support Team	☐
13.2 Design, develop and implement a database for storing the subjective data collected from participants (e. g., from questionnaires, from focus groups, from feedback lines etc)	Project Manager, Research Team, Technical Support Team	☐
13.3 Develop data navigation and visualization tools	Project Manager, Research Team, Technical Support Team	☐
13.4 Sign off on database for storing FOT data.	Project Manager, Research Team, Technical Support Team, Project Management Team, Legal and Ethical Advisors, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ Before an FOT is launched, the database architecture should be reviewed by a system evaluator to ensure that all requirements are fulfilled.
- ✓ *Ensure copies are made of raw data, reduced raw data and all processed data files and store these securely, separate from the primary data store.*
- ✓ Use an industry standard relational database to store the data.
- ✓ *Ensure that unauthorised access to the database is not possible. Preferably, do not give the database host an IP number.*
- ✓ Careful database design can reduce the need for post-collection manipulation if the database is designed to feed directly into a statistical package for data cleaning and analysis.
- ✓ *Decide early in the project how to manage post-project data. Issues to consider are: What happens to data when the project ends? Who will have data usage rights? Who can access it? Who pays for possible storage? In projects with large amounts of stored data (several terabytes), the cost to store and manage data is not insignificant, and all project partners might not have the means to handle it afterwards. Where data is taken off-line, determine what meta data should be kept, and how.*

General Advice

- ✓ See chapters **Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.** of the FESTA handbook for more detailed advice relating to this activity.
- ✓ Basic legal advice on this issue is also provided in Deliverable D6.3 and the **Errore. L'origine riferimento non è stata trovata.**and FOT-Net 2, D3.2

Activity 14: ***Test all systems against functional requirements and performance specifications***		
Tasks and Sub-Tasks	**Person/ Team/ Organisation Responsible for Activity**	**Done**
14.1 Develop "acceptance testing" protocols (see comment column).	Project Manager, Research Team, Technical Support Team	☐
14.2 Test the systems for acceptance, using the acceptance testing protocol	Project Manager, Research Team, Technical Support Team	☐
14.3 Develop a usability test plan for the purpose of assessing the systems and functions for usability	Project Manager, Research Team with consultant (if appropriate)	☐
14.4 Conduct usability testing, using the usability testing plan, to ensure systems and functions are user-friendly and that they meet all usability assessment criteria	Project Manager, Research Team, with consultant (if appropriate)	☐
14.5 Obtain or develop a valid and reliable ergonomic checklist	Project Manager, Research Team	☐
14.6 Assess systems, using the ergonomic checklist, to ensure that they meet all relevant criteria.	Project Manager, Research Team	☐
14.7 Assess vehicles against relevant certification procedures to ensure that vehicles are safe, roadworthy and comply with all relevant National, State and Territory laws, treaties and other protocols.	Project Manager, Technical Support Team with consultant (if appropriate)	☐
14.8 Ensure that all vehicle modifications that affect primary safety are signed off by a competent engineer or appropriate testing authority.	Project Manager, Technical Support Team with consultant (if appropriate)	☐
14.9 Rectify all technical, usability, ergonomic and certification issues where deficiencies are noted.	Project Manager, Research Team, Technical Support Team with consultant (if appropriate)	☐
14.10 Sign off on completion of all systems tests	Project Manager, Research Team, Technical Support Team, Project Management Team, Project Sponsor(s)	☐
Critical Considerations (the "dos" and "don'ts")		

- ✓ *Do not sign off on the outputs of any of the previous activities until all technologies have been tested and, where appropriate, refined.*
- ✓ Be sure that all systems are designed so they do not drain the battery when the engine is not running.
- ✓ Be sure that retrofitted systems are properly secured and meet all relevant crashworthiness requirements.
- ✓ If sub-contractors are appointed to install or maintain test equipment, implement a quality assurance programme.
- ✓ *Be aware that system clocks can drift significantly if left to run independently . Where feasible, use GPS time to correct system clock error.*
- ✓ *Implement procedures to ensure that alignment and calibration of sensors is maintained and tested in all potential weather conditions.*
- ✓ Various guidelines, standards and checklists exist for assessing the ergonomic quality of the human-machine interface for ICT systems (see Regan, Lee and Young, 2008, for a summary). Also see: AIDE (EU-Project) Deliverable 4.3.1: "Report on the review of available guidelines and standards" – publicly available over the internet.
- ✓ *Be aware that some system components may become corrupted over time with continuous use (e. g. flash memory cards).*
- ✓ Revisit the installation manual for all platforms.
- ✓ Consider the need to obtain waivers/special licences from regulatory authorities for equipment that is non-compliant (e. g. radars that operate outside legal bandwidths).
- ✓ Standard testing of vehicle modifications by a competent authority may be necessary with respect to safety features (e. g. proper deployment of airbags following modification to vehicle interiors).
- ✓ *Be aware that some systems (e. g. displays) that are not OEM-installed may fail in automotive environments.*
- ✓ *Where appropriate, test for radio frequency (RF) interference effects (e. g. from overhead tram wires), which may adversely affect system operation. Also ensure that normal vehicle systems (e. g. FM radio and remote locking) are not affected by installed equipment.*
- ✓ Ensure that the computers powering the data collection system and support systems are powerful enough to ensure that the data sampling rate is consistent and at the rate specified.
- ✓ Don't assume that OEM systems that are already installed in test vehicles have been ergonomically assessed against appropriate standards and guidelines. Ergonomic assessment of systems prior to system deployment can be useful in identifying ergonomic problems that may explain or confound treatment effects.
- ✓ Provide a written statement for the participants to keep (in the vehicle) which confirms their participation in the FOT and the nature of vehicle modifications – in case they are challenged by Police or other authorities.
- ✓ *Resolving any technical, usability, ergonomic, and certification issues may require several iterations. Do not underestimate the time required for this process.*

General Advice

- ✓ This activity is *not* about pilot testing — it is about testing the performance, security and reliability of systems – to ensure that all technologies to be deployed perform in accordance with the functional requirements and performance specifications developed for them in previous activities.
- ✓ An Acceptance testing Protocol is a test protocol for testing that all systems to be used in the study (FOT systems, data collection systems and support systems) meet the functional requirements and performance specifications developed for them by the FOT project team, under all foreseeable operating conditions.
- ✓ The term "usability" can mean different things to different people. The test plan should use a standard definition of usability (e. g. ISO 9241).
- ✓ Be aware that the frequency used by some radar-based systems may interfere with the operation of other systems used by Police, emergency services or other operators (or vice versa) when used in other countries or jurisdictions. This must be investigated where the FOT is conducted across State and international boundaries.

Activity 15:

Develop FOT recruitment strategy and materials

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
15.1 Develop recruitment strategy, including user entry and exit requirements and procedures.	Project Manager, Research Team, Legal and Ethical Advisors	☐
15.2 Develop recruitment materials and procedures	Project Manager, Research Team, Public Relations and Communications advisor	☐
15.3 Sign off on recruitment strategy, materials and procedures.	Project Manager, Research Team, Project Management Team	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ Consider whether participants should be representative of the relevant population to ensure generalisability of results.
- ✓ Assume that there will be an attrition rate of about 10 to 15 % when using company employees, who come and go, and retire.
- ✓ Be aware that, when company employees change jobs within their companies, this may have a dramatic effect on their annual travel.
- ✓ If fleet drivers are recruited via a fleet owner or manager it is also necessary to get buy-in from individual drivers.
- ✓ With respect to safety, select drivers who do not pose a risk to themselves, others or the project. Be aware of the potential for bias in the results.
- ✓ Do not underestimate the complexities involved in recruiting company employees.
- ✓ Be aware that some commercial operations may have employee turn-over rates approaching 100 % per annum.

General Advice

- ✓ See chapters **Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.** for further advice relevant to this Activity.
- ✓ *The Ethical requirements for recruitment of users may be difficult to adhere to when recruiting company employees.*
- ✓ Ideal companies to approach to recruit fleet vehicle drivers have the following characteristics: many vehicles; drivers have high mileage rates; drivers drive primarily in the geographical areas of interest in the FOT; and management has a commitment to the aims and objectives of the FOT.
- ✓ It is not possible in many countries to obtain personal information about drivers that can be used to screen them for inclusion in the study (e. g., has a drunk driving record).
- ✓ It may not be possible in some countries to obtain directly from car dealers the names of drivers of particular makes and models of vehicles.
- ✓ In some countries (e. g., France), potential participants must be screened by a registered doctor.
- ✓ *The recruitment materials and procedures will need to have been incorporated and approved as part of the FOT ethics and legal approval processes.*

Activity 16:

Develop training and briefing materials

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done

16.1 Conduct training needs analysis (TNA) to identify training requirements of participants and other relevant actors	Project Manager, Research Team with consultant (if appropriate)	☐
16.2 Design and develop briefing and training materials, based on outputs of the TNA	Project Manager, Research Team with consultant (if appropriate)	☐
16.3 Design and develop FOT system(s) user manual (if appropriate)	Project Manager, Research Team, Legal and Ethical Advisors	☐
16.4 Design and document the procedures for the delivery of the briefing and training to the FOT participants	Project Manager, Research Team	☐
16.5 Sign off on training and driver (and company) briefing materials and delivery processes.	Project Manager, Research Team, Project Management Team	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ *Ensure that training programs and briefing materials are designed in a way that does not confound experimental treatment effects.*
- ✓ Ensure all users understand existing all systems and functions to be used (including test systems).
- ✓ Don't underestimate the time required for the development of briefing and training materials — it is a time consuming activity.
- ✓ When pre-testing the user-friendliness of a function a self-learning approach may be used.
- ✓ Be aware that an excess of training might affect the possibility to understand the short-term unintended effects of the system.
- ✓ Provide drivers with a mini-operating manual to keep in the vehicle and prepare written materials (brochures, DVDs & CDs) that can be taken away after briefing sessions.

General Advice

- ✓ See chapters**Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.** for further advice relevant to this Activity
- ✓ See Regan et al, 2006 (Volume 2) for examples of training and briefing materials used in a previous FOT.
- ✓ *Refresher training may be required if FOT systems are not activated for several weeks or months into the FOT.*
- ✓ The training and briefing materials and procedures will need to have been incorporated and approved as part of the FOT ethics and legal approval processes.

Activity 17:

Pilot test FOT equipment, methods and procedures

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
17.1 Develop protocol for pilot testing FOT equipment, methods, procedures, evaluation tools and materials (including training, briefing materials and data collection, downloading and analysis procedures)	Project Manager, Research Team, Technical Support Team, Legal and Ethical Advisors	☐

17.2 Recruit, brief and train pilot participants	Project Manager, Research Team	☐
17.3 Deploy a small sample of FOT platforms under a representative range of external conditions that will be experienced in the FOT, as per the pilot testing protocol.	Project Manager, Research Team, Technical Support Team	☐
17.4 Fine tune FOT platforms and technologies, systems, procedures, evaluation tools and protocols, as required, on the basis of the pilot data yielded.	Project Manager, Research Team, Technical Support Team, Project Management Team	☐
17.5 Sign off on pilot testing.	Project Manager, Research Team, Technical Support Team, Project Management Team	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ *Do not truncate your pilot test plan, and do not underestimate the time required for comprehensive pilot testing. The importance of pilot testing cannot be overstated.*
- ✓ Undertake a 'full dress rehearsal' with participant involvement and a duration that is representative of the duration that will occur in the FOT.
- ✓ Use pilot testing also as a means of estimating the amount of time required to complete activities, as this will enable more accurate budgeting during the remainder of the project.
- ✓ Pre-test all data analysis procedures to ensure appropriate data is collected – particularly data related to event recording triggers.
- ✓ Ensure that the routes used in pilot studies maximise the likelihood of critical situations of relevance to the FOT. Consider using a test track to verify the logging of critical situations.
- ✓ Add independent monitoring systems to pilot platforms to ensure the validity of data derived from sensors.
- ✓ *In the pilot phase listen to the users and, when involved, owners and managers of the vehicle fleet – their ideas are likely to be different.*

General Advice

- ✓ See chapter **Errore. L'origine riferimento non è stata trovata.**, **Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.** for further advice relevant to this Activity.
- ✓ For data collection systems, ensure that data is being recorded, determine the accuracy of data recorded, test downloading procedures and equipment, test reader software and analyse samples of pilot data.

Activity 18:

Run the FOT

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
18.1 Ensure that all sign offs have occurred for previous activities.	Project Manager, Project Management Team	☐

18.2 Manage the FOT: • **monitor project activities, timelines, budgets and resources** • **prepare regular progress and financial reports for sponsor** • **convene and attend regular meetings with research and support teams** • **maintain communication with sponsor and key stakeholders**	Project Manager, Research Team, Technical Support Team, Administrative Support Team, Project Management Team	☐
18.3 Recruit participants	Project Manager, Research Team	☐
18.4 Organise training session times/materials	Project Manager, Research Team	☐
18.5 Brief and train participants	Project Manager, Research Team	☐
18.6 Brief fleet managers (if appropriate)	Project Manager, Research Team	☐
18.7 Deploy FOT platforms	Project Manager, Research Team, Technical Support Team	☐
18.8 Regularly monitor participant progress, including kilometres travelled	Project Manager, Research Team	☐
18.9 Administer questionnaires and implement other data collection methods at pre-determined intervals	Project Manager, Research Team	☐
18.10 Collect, enter into database (unless automated) and store subjective data	Project Manager, Research Team	☐
18.11 Record, download and store objective (i.e., logged) data	Project Manager, Research Team, Technical Support Team	☐
18.12 Collect special data (e. g., fuel dockets) needed to analyse surrogate performance indicators	Project Manager, Research Team	☐
18.13 Monitor for, collect and document data on technical problems and user feedback	Project Manager, Research Team, Technical Support Team	☐
18.14 Commence preliminary evaluation of data, to identify instances of dangerous driving and any other findings of interest/relevance to FOT outcomes	Project Manager, Research Team, Technical Support Team	☐
18.15 Repair and re-deploy platforms (as required)	Project Manager, Technical Support Team	☐
18.16 Routinely ensure all platforms are properly maintained and legal in other ways	Project Manager, Research Team, Technical Support Team	☐

(e. g., registered, licensed, tyres properly inflated)		
18.17 Report dangerous driving behaviours (if legally required)	Project Manager, Research Team, Technical Support Team	☐
18.18 Conduct exit interviews with users and the other relevant actors	Project Manager, Research Team	☐
18.19 Remove systems and equipment from private vehicles (if used)	Project Manager, Technical Support Team	☐
18.20 Sign off on completion of this activity of the FOT.	Project Manager, Research Team, Technical Support Team, Administrative Support Team, Project Steering Committee, Project Management Team, Accounting/Auditing Advisor, Sub-Contractors	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ Anticipate, and plan for, participant 'dropout' throughout the FOT — over-sample. It is rarely possible to replace participants who drop out after more than a few days without affecting the timing plan.
- ✓ *Develop protocols for responding to drivers with technical and other problems (e. g., provide drivers with a dedicated cell phone to report problems; ensure at least two people have pagers to receive problem calls; etc) Timely responses will keep drivers happy.*
- ✓ Anticipate problems that may increase the drop out rate (e. g., higher fuel consumption in the FOT vehicle than in the drivers' own vehicle) and take steps to prevent or mitigate these problems.
- ✓ Monitor closely system usage for drivers who may be tempted to 'demonstrate' novel systems to friends and neighbours.
- ✓ Adhere to quality control mechanisms to ensure that data is being properly recorded and downloaded.
- ✓ Adhere to calibration procedures to ensure accuracy of measurements/sensors over time and help prevent data drift issues.
- ✓ Find a suitable location for training drivers where you can also assess transfer of training to the test vehicles in a safe environment
- ✓ *If the number of kilometres driven by drivers is being controlled for, conduct regular calibration checks of cumulative distance travelled.*
- ✓ Assume that it will take you 50 % longer than you think to recruit participants if recruiting company drivers.
- ✓ Check logged data as soon as you receive it to verify accuracy and completeness of data and verify kilometres travelled.
- ✓ *Monitor and record critical factors that could have an impact on the measured outcomes/dependent variables (e. g., changes in Police enforcement strategies, unseasonal weather conditions). If these are not controlled for in the experimental design, or accounted for in the analyses, they could confound the measured effects of the systems being tested.*
- ✓ Where company fleet vehicles are involved in the study, advise fleet managers not to "demonstrate" their vehicles, as this may compromise the aims of the study.
- ✓ *Give sponsors early warning of potential problems that could compromise the integrity of the study, or increase the budget.*
- ✓ Encourage participants to report technical problems as soon as possible.
- ✓ *Don't assume that all systems in the test vehicles are functioning as required. Develop systems to check, at appropriate times, that they are operating properly.*
- ✓ Don't assume that drivers will do what you ask them to do (e. g., to fill out questionnaires; maintain vehicles). They need regular reminding and follow-up.
- ✓ Where data downloading is manual, don't forget to replace flash memory cards, or other storage devices, with new (empty) ones on a regular basis.
- ✓ Do not always assume that drivers will clock up their kilometres evenly over the trial. Contact them on a regular basis to check cumulative distance logged.
- ✓ If legally required, don't forget to report to the appropriate authorities (e. g., company fleet managers) recorded instances of dangerous driving by test drivers.
- ✓ *Don't assume that drivers will drive the vehicles without trailers, bike racks and other accessories. These may affect the operation of some FOT systems (e. g., reverse collision*

warning devices).

- ✓ Minimise interference to commercial operations during FOTS, especially trucking operations. Problems that compromise commercial productivity may result in companies withdrawing trucks from the FOT.
- ✓ *Make sure fleet managers are, and remain, motivated. Their support is critical.*
- ✓ Be careful about the feedback given to drivers. They may be concerned about the possibilities of 'unintended consequences' e. g., their managers learning how and when they take rest breaks etc.
- ✓ Participants are more likely to comply with what is asked of them if they engage with the project. Ongoing communication and even small incentives can enhance perceived engagement and improve compliance. However, the level of engagement must not compromise the outcomes of the study.
- ✓ Remember that long-term involvement in a research study can be onerous for a participant. At all times treat them as participants in the study process, not simply subjects of a study.
- ✓ *Allow sufficient time for any data entry which has to be done manually (e. g., responses from pencil and paper questionnaires, focus groups). As far as is possible, manual data entry should be carried out routinely during the course of the data collection phase and not all left to the end.*
- ✓ A system for basic inventory management is recommended for FOTs with more than a few vehicles in use. For such a system to be efficient, sensors, data acquisition system units, vehicles and all other equipment need to be included, as well as relevant supporting procedures developed.

General Advice

- ✓ More detailed advice can be found in other chapters of this Handbook: Chapter **Errore. L'origine riferimento non è stata trovata.** for participant recruitment; Chapters **Errore. L'origine riferimento non è stata trovata.** and Chapter **Errore. L'origine riferimento non è stata trovata.** for organising training sessions; Chapters **Errore. L'origine riferimento non è stata trovata.**, **Errore. L'origine riferimento non è stata trovata.** for implementing data collection methods at pre-determined intervals and for collecting and storing subjective data; **Errore. L'origine riferimento non è stata trovata.** for preliminary data analysis; Chapter **Errore. L'origine riferimento non è stata trovata.**, **Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.** for vehicle maintenance and compliance with laws; and Chapter **Errore. L'origine riferimento non è stata trovata.** for reporting of dangerous driving, where appropriate.
- ✓ *Ongoing communication with key stakeholders is important during the FOT to ensure that the aims and objectives of the FOT are clear, that stakeholders stay committed to the project, and that the aims and objectives of the FOT are not misquoted, misrepresented or misunderstood.*

Activity 19:

Analyse FOT data

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
19.1 Develop a data analysis plan	Project Manager, Research Team, Technical Support Team	☐
19.2 Analyse objective (i.e., logged and recorded data)	Project Manager, Research Team	☐
19.3 Analyse subjective data (i.e., data obtained from interviews, questionnaires, focus groups, hotlines, etc)	Project Manager, Research Team	☐

19.4 Draw conclusions with respect to the hypotheses generated for the FOT	Project Manager, Research Team	☐
19.5 Sign off on completion of all required analyses	Project Manager, Research Team, Project Management Team, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ Plan for the fact that there will be constant demand for study findings, such as general trends in the data, early in the project, even though the data may not be statistically reliable enough to report with any confidence.
- ✓ In a well-powered study, null findings (i.e., where no effect is found and the hypotheses refuted) are potentially as interesting as when the hypotheses are supported.
- ✓ *Anticipate the requirement to have to perform supplementary analyses for the funding organisation, which may be expensive and not originally budgeted for. This will require negotiation with the sponsor if these analyses are expected to be carried out within the original budget.*
- ✓ Anticipate that, unless distance travelled is controlled for in the FOT, the distance travelled by different drivers will vary significantly. Take this into account in the analysis to ensure results are not skewed.
- ✓ Don't forget to run "reality checks" on the data, to be sure that the data are "clean". This is essential.
- ✓ *If data is reduced/aggregated, always keep a copy of un-aggregated data.*
- ✓ Ensure that all data analysts have used the test vehicles and understand the circumstances in which data was/is collected.
- ✓ All team members who handle participant data should receive appropriate training regarding data privacy.
- ✓ Work out how to best filter logged data and deal with missing data.

General Advice

- ✓ See Chapters **Errore. L'origine riferimento non è stata trovata.**, **Errore. L'origine riferimento non è stata trovata.** and **Errore. L'origine riferimento non è stata trovata.** for detailed advice on data analysis tools and methods.
- ✓ There may be a requirement to conduct ongoing analysis, such as ongoing identification of dangerous drivers, determining whether adaptation to systems is occurring early enough to warrant a shorter FOT duration (e. g., to save money and time), and to identify early trends in the data. These checks should be built into the analysis plan at the start of the project.
- ✓ Some FOTs have developed novel ways of turning ADAS technologies on and off to control precisely the amount of exposure to the technologies that are being evaluated (see Ref 1).
- ✓ Sponsors need to be calibrated about the relative costs of running FOTs. For example, the cost of running simulation models at the end of the FOT to estimate safety and other benefits of ICT technologies is a fraction of the cost of preparing and deploying the FOT vehicles.

Activity 20:

Write minutes and reports

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done

20.1 Write minutes of regular project management team meetings	Project Manager	☐
20.2 Write regular minutes of Project Steering Committee meetings	Project Manager	☐
20.3 Write quarterly progress reports for the sponsor(s)	Project Manager	☐
20.4 Write the draft FOT report	Project Manager, Research Team, Technical Support Team	☐
20.5 Send the draft FOT report to relevant stakeholders and peers for peer-review	Project Manager	☐
20.6 Convene 1 or 2 meetings to discuss feedback with sponsor/peers	Project Manager	☐
20.7 Incorporate feedback and write final report.	Project Manager, Research Team, Technical Support Team	☐
20.8 Deliver final report to sponsor(s)	Project Manager	☐
20.9 Sign off on completion of all required reports	Project Manager, Research Team, Technical Support Team, Project Management Team, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ *Use regular progress reports to document problems, solutions and lessons learned.*
- ✓ Allow sufficient time for sponsor review of draft and final reports, but not so long that the review process drags out unduly. Six to 8 weeks is recommended.
- ✓ Consider peer review of major outputs; this will improve their quality but delay their release.
- ✓ *Document all lessons learnt in the final FOT report.*
- ✓ Ensure that the final report contains practical recommendations for wider scale deployment of those systems found to be effective, and for fine-tuning of those with potential to be more effective.
- ✓ *Develop, in consultation with the Project Steering Committee, a suggested plan for implementing the recommendations deriving from the FOT. Document the implementation plan in the FOT final report.*

General Advice

- ✓ *The FOT lifecycle is long. Hence, it is advisable to write separate reports on each critical stage of the FOT, particularly the lessons learned, to ensure that nothing important that should be documented is forgotten.*
- ✓ Formal meeting minutes are a critical resource for the project in confirming departures from the project plan.

Activity 21:

Disseminate the FOT findings

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done
21.1 Send regular project reports to the sponsor	Project Manager	☐
21.2 Disseminate preliminary and final findings at seminars, conferences and special events	Project Manager, Research Team, Technical Support Team	☐
21.3 Prepare reports on preliminary findings for the sponsor	Project Manager, Research Team, Technical Support Team	☐
21.4 Send sponsor draft and final FOT reports	Project Manager	☐
21.5 Provide other stakeholders with access to FOT final report (s) and, if allowed, raw or filtered data from the FOT	Project Manager, Research Team, Technical Support Team	☐
21.6 Showcase the vehicles at relevant events during the FOT (e. g., Smart Demos, motor shows) to promote awareness and wider deployment of systems.	Project Manager, Technical Support Team, Project Steering Committee, Project Management Team, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ Disseminate the findings in accordance with the previously agreed communications plan.
- ✓ *Agree on what can and cannot be disseminated and said at different points in the study.*
- ✓ Seek necessary permissions prior to divulging FOT findings to any third party.
- ✓ FOT reports are large and expensive to print. Allocate sufficient budget at the beginning of the project for printing.
- ✓ FOT reports are large and hard to read. It is desirable to produce conference papers along the way that document the outputs of the study at different phases. Prepare a concise 1 or 2 page synopsis of the study outcomes that can be read and easily digested by politicians, chief executives and relevant others in positions of authority.
- ✓ Agree in advance who is empowered to release and comment on results.

General Advice

- ✓ Where private industry is a participant in the FOT, it may be necessary to seek permission from the manufacturer before divulging certain information deriving from the FOT. This must be established.
- ✓ Maintain at least one vehicle for demonstrations; preferably at a location that is convenient to politicians, officials and the press.
- ✓ A demonstration and briefing to an influential politician is likely to be far more effective than sending them a report.

Activity 22:

Decommission the FOT

Tasks and Sub-Tasks	Person/ Team/ Organisation Responsible for Activity	Done

22.1 Conduct de-briefing interviews with participants to elicit feedback on the FOT that can be used to improve future FOTs.	Project Manager, Research Team	☐
22.2 Dispose of test vehicles which are no longer needed (if vehicles are not privately owned).	Project Manager, Research Team, Technical Support Team, Administrative Support Team, Project Management Team, Accounting/Auditing Advisor, Project Sponsor(s)	☐
22.3 Retrieve installed data logging equipment (if vehicles are privately owned)	Project Manager, Research Team, Technical Support Team, Administrative Support Team, Project Management Team, Accounting/Auditing Advisor, Project Sponsor(s)	☐

Critical Considerations (the "dos" and "don'ts")

- ✓ Ensure that participants return relevant items at the end of the study (e. g., flash memory cards, i-buttons) and perform other required activities to decommission the FOT vehicles (e. g., disconnect power to support systems).
- ✓ *Keep one vehicle until all data analyses are complete.*
- ✓ Consider providing public access to FOT databases, where ethically allowed, that enables others to use the data for other research purposes after the FOT has been de-commissioned (but remember to fully explore and address anonymity issues). The data collected and stored after the FOT is de-commissioned should be regarded as "living data".
- ✓ *Don't lose momentum at the end of the FOT. Lobby stakeholders to ensure that there is commitment to implementing the recommendations of the FOT.*

General Advice

- ✓ *Consider keeping one or two vehicles as showcasing vehicles after the study, to allow stakeholders in positions of authority to experience the look and feel of the vehicles.*
- ✓ It may be necessary to consider legal issues of decommissioning the FOT as far as the de-installation of data logging equipment is concerned (in a contract with participants). Deliverable 6.3 and Annex A and FOT-Net 2, D3.2 on legal issues.

FESTA Handbook

Version 5

Annex C Defining crash relevant events in NDS/FOT studies formulation

---

Revised by FOT-NET

(Field Operational Tests Networking and Methodology Promotion)

28 February 2014

ABA	Aggregation Based Analysis
ACC	Adaptive Cruise Control
ACEA	European Automobile Manufacturers' Association
ADAS	Advanced Driver Assistance Systems
CAN	Controller Area Network
CRE	Crash Relevant Events
EBA	Event Based Analysis
EuroNCAP	European New Car Assessment Programme
FCW	Forward Collision Warning
FESTA	Field opErational teSt supporT Action
FOT	Field Operational Test
GMM	Gaussian Mixture Models
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GSM	Global System for Mobile communications
HMI	Human Machine Interface
I2V	Infrastructure to Vehicle(see Vehicular communication systems)
ICT	Information Communication Technology
ISA	Intelligent Speed Adaptation
ITS	Intelligent Transportation System
LDW	Lane Departure Warning
NDD	Naturalistic Driving Data
NDS	Naturalistic Driving Study
NHTSA	National_Highway_Traffic_Safety_Administration
OEM	Original Equipment Manufacturer (in this document used as synonym of vehicle manufacturers)
RR	Range Rate
TTC	Time To Collision (relevant parameter in preventive safety applications – see e.g. Prevent)
V2I	Vehicle to Infrastructure (see Vehicular communication systems)
V2V	Vehicle to Vehicle (see Vehicular communication systems)
V2X	Vehicle to X (Any Station) (see Vehicular communication systems)
VMS	Variable Message Sign
WYLFIWYF	What-You-Look-For-Is-What-You-Find

Annex C - Defining crash relevant events in NDS/FOT studies formulation

1. Introduction

Non-intrusive logging and analysis of real driver behaviour in real traffic comes in two flavours, i.e. Naturalistic Driving Studies (NDS) and Field Operational Tests (FOT). For NDS, the key goal is usually to gain an understanding of crash causation mechanisms and in particular, which driver behaviours are associated with increased crash risk. For Field Operational Tests (FOT), the key goal usually is to evaluate whether one or more (usually in-vehicle) safety technologies has a detectable and significant influence on crash risk (be it positive of negative).

In NDS/FOT studies, massive amounts of vehicle and video data are usually collected on a continuous basis. Once processed and uploaded, this generates an enormous amount of data for analysis in the database. While this data obviously can be sliced and sorted in very many ways and for many purposes, the focus of this report lies on how to approach the analysis in a way that leads to an understanding of how driver behaviours and/or vehicle systems influence crash risk. To contrast this with other types of analysis, one example would be using the data to identify natural acceleration and deceleration patterns that can be used for fuel economy tuning of vehicle engines.

For NDS/FOT data, there are two general analysis approaches that can be taken toward uncovering behaviours and/or system influences that drive changes in crash risk. The first can be called Aggregation Based Analysis (ABA). The basic principle of ABA is to identify trends and/or patterns in performance measures that have been aggregated over longer time segments. A typical example of such an analysis would be whether average time headway or average travel speed changes when drivers use Cruise Control. If there is a significant change, then that change might in a next stage be used to predict possible changes in crash/injury/fatality risk if the effects were to be extrapolated to the general driver population. However, in this report, ABA types of analysis will not be addressed. The reason for this is that such extrapolations are very difficult to justify, due to the fact that the connection between average measures and crash risk is hard to establish. This will be further discussed below. The other analysis approach is Event Based Analysis (EBA). The basic principle of EBA is to identify shorter driving segments (typically in the order of 5-10 seconds), during which the risk of crashing is judged to be higher compared to other driving in the data set, and then to analyse these events further. These events are often referred to as Crash Relevant Events (CRE), since their occurrence is thought to be indicative of actual crash risk in one way or another.

In NDS, analysis of CREs usually focuses on establishing why they occur from a driver perspective. In particular, the aim is to find out whether some types of driver behaviour are overrepresented in CREs as compared to baseline events, i.e. shorter driving segments where crash risk is judged not to be elevated. If particular behaviours can be identified as occurring disproportionally often during CREs compared to baseline events, then it is generally assumed that they can be viewed as contributing to the elevated crash risk, which makes them targets for countermeasure development. In FOTs, the focus when analysing CREs is somewhat different. One key question is whether the frequency of various CRE types goes up or down as a function of Advanced Driver Assistance Systems (ADAS), i.e. if drivers

experience critical situations less (or more) often when using ADAS. The other key question is whether driver responses during the CREs that occur is different when ADAS are being used. For example, does accelerator release come earlier when drivers are given a collision warning; do they brake harder when warned, etc.…). These types of analysis are not the same as looking for risk increasing behaviours, something which is important to keep in mind when setting up an FOT study.

2. Defining Crash Relevant Events

From the above, it follows that a key element to NDS/FOT success is defining CREs in a proper way. If the events selected for analysis are indeed crash relevant, then extrapolation to the general driver population is indeed both possible and credible.
However, while this is fine in theory, identifying CREs in NDS/FOT data is a bit more difficult in practice. To begin with, the simplest and most direct measure of crash risk, i.e. actual crashes, are incredibly rare events. Even if hundreds of drivers are being continuously observed during several years of driving, the statistically low likelihood of having an actual crash means that the number of crashes in the final database will be smaller than required for statistical analysis, even if the database itself is huge.

It follows that surrogate events have to be used. These events have to have very particular properties, i.e. they need to be critical situations where there is no actual crash, but where the event still unfolds in such a way that its presence can be used as an indicator of crash risk.

The iceberg ratio metaphor is typically used here. For example, if one assumes that there are 100 incidents for every accident in a certain working place, then a measured reduction of incidents by 50% could be used to predict a corresponding 50% accident reduction in the future, even if there's been no accident yet. A key challenge to analysis and interpretation of NDS/FOT data is therefore how to couple non-crash events to crash causation mechanisms. In principle, if the link between crash causation and a CRE type is weak, then any interpretation of the CREs frequency or behavioural content is correspondingly weak, and vice versa.

Ideally, one would therefore only use CRE types that are known with certainty to be predictive of actual crash involvement, i.e. for which it is legitimate to infer that a change in their frequency corresponds to a (proportional) change in crash risk. Unfortunately, precise CRE definitions with clear cut, undisputable connection to crash involvement have yet to be fully established (had they done so, a simple list below would have concluded this report nicely).

To illustrate; while hard braking events might seem a plausible candidate, in the VTTI 100 car study (Dingus, Klauer et al. 2006) it was not possible to reliably identify critical driving events based on hard braking alone, since hard braking also occurred in many situations not associated with elevated crash risk.

2.1 Approaches to Crash Relevant Event Detection

This problem of identifying relevant CREs is not new. A lot of effort in many projects has gone into developing algorithms, filtering techniques etc. that allow for efficient yet relevant CRE selection. The aim here is not to list each such CRE in detail. Instead, the aim is to

describe the approaches to data analysis that they represent on a higher, grouped level, and then go through some of the pros and cons that each CRE group, or analysis approach, faces. Selecting among these approaches to CRE definition is a major decision point for a project. Projects must make an informed and conscious decision as to which of the below approaches will best fulfil their goals, and then set up CREs correspondingly. There are pros and cons to each approach, some of which may have large impact on the results.

2.2 Approach 1 - Driver response based identification

The first approach can be called the “driver response based” approach, and it builds on the general idea that CREs can be identified from the way drivers respond to them.

The most common version of this is to look for extreme vehicle kinematics. The basic assumption is that drivers prefer to travel in comfort and generally will not expose themselves to kinematically drastic events unless necessary. Thus abrupt velocity and direction changes in the vehicle (e.g. hard accelerations /decelerations and/or rapid steering) or abrupt movements of the driver (are thus considered to be out of the ordinary, indicating an unplanned and urgent response to an unexpected situation (of course, this might capture “play” with the vehicle as well).

There are two main ways to identify such drastic manoeuvres. Principally, one can either look for momentary or sustained breaches of defined thresholds. Looking for momentary breaches means identifying all instances where a particular threshold is exceeded, regardless of the excess duration. As an example, one could look for all instances of Brake Pressure Jerk (BPJ), above a certain value.

Just looking at the momentary threshold breaches might however capture many false positives. Kinematic signals can have high momentary peaks within their normal operation interval (e.g. acceleration spikes due to potholes in the road), and vehicle sensors are more often than not rather “noisy”, i.e. signal variation does not correspond to real variations in the parameter. To remedy this (apart from signal filtering, which is more considered signal pre-processing before database upload and hence not covered here), an often used approach is to add a minimum time during which the threshold has to be breached.

For example, for deceleration levels above 0.8G, one can add a criterion that it needs to stay above 0.8G for more than 0.5 seconds for the event to count as a CRE. Of course, if the restrictions put in place are too strict, no CREs will be found in the data. Thus it is always necessary to strike a balance between removing false positives and keeping true positives.

A less common, but potentially promising approach, is to use what can be called the startle response in the driver to find events. The general idea is that unexpected traffic situations that also include a perceivable threat to the driver triggers a response in the form of an adrenalin rush and a general tensioning of the body. This bodily “jerk” may be used as a tell-tale that the driver did not expect the situation and that s/he perceives it as genuinely threatening, even though s/he may not brake or steer the vehicle in a particularly dramatic way.

2.3 Approach 2 - Function response based identification

If the study is of the FOT kind, i.e. designed to assess the impact of one or more active safety functions, then a very natural approach to CRE identification is to use the function itself to detect CREs. After all, that is what the function is designed to do.

For example, if an FOT is set up to assess the effects of Forward Collision Warning (FCW) on crash risk, the warnings issued by the collision warning system (even though they are not shown to the driver in the baseline phase) can be used as event identifiers.

The downside to this approach is of course that any CRE that occurs outside the function's detection capacity or that occurs when the system is turned off (in the treatment phase) will be missing from the analysis. Relying on the system signals only thus makes it impossible to estimate the frequency of CREs which the function in principle needs to detect but in practice cannot, though their prevention would enhance traffic safety.

Looking at the other side of the coin, the advantage of the approach is that it is not a worry how to factor in function availability and usage in the safety analysis. Since the function only can do something when it is turned on and does detect a threat, warnings/interventions can only occur when the function is capable of delivering them. Hence the true availability and usage rates are automatically represented in the data set.

2.4 Approach 3 - Driving context based identification

A third approach to CRE identification is to avoid relying on driver or function responses and instead make it driving context based. The underlying assumption of this approach is that too small margins equal elevated crash risk. In other words, there exist situations where the safety margins are inherently so small that the slightest mistake or variation could lead to a crash. Since mistakes do occur, it follows that crashes will also occur under these vehicle and/traffic environment configurations, and prevention of whatever it is that leads to these small margins thus will enhance traffic safety.

The definition of what constitutes too small margins can be either static or dynamic. An example of a static approach is the one used in the Road Departure Crash Warning System Field Operational Test (LeBlanc, Gordon et al. 2006). When evaluating the influence of a Lane Departure Warning system, the researchers used the lane marker as a static boundary, i.e. something that by definition should not be crossed unless the crossing is intentional.

Following this definition they looked for all instances of driving where the vehicle was travelling at or above the speed where the system would be available and where the vehicle inadvertently left the lane (defined as crossing the lane marker without having the turn signals activated).

An example of a dynamic version of the same approach is Najm et al. (Najm, Stearns et al. 2006), who developed a CRE algorithm that defined too small margins as being within a certain kinematic envelope. CREs were selected by identifying all situations where the vehicles fell below a combined Time-To-Collision (TTC) and Range Rate (RR) threshold in relation to a lead vehicle. The basic idea was that if a vehicle is closer than X to another

vehicle and simultaneously closing in faster than Y, then there is a conflict regardless of whether an action is taken or not.

Four levels of conflict intensity were defined, depending on the precise TTC and RR thresholds. The values for Y as a function of X were selected based on empirical data from normal and deliberately delayed driver responses on a test track in lead vehicle following situations (Kiefer, Shulman et al. 2003). In other words, they essentially picked a demarcation line which says that if a vehicle is in this dynamic situation, the driver is by definition in trouble since most drivers would already have braked by now.

The advantage of this approach is that the actual conflict definition is highly objective, in the sense that it is based on vehicle kinematics only and does not depend on how the driver responds to it. Thus, while the setting of the boundaries for the kinematic envelope that defines the conflict might take some discussion, there will not be later disagreement as to whether the event was crash-relevant or not.

However, this also constitutes the weakness of the approach. Most importantly, active drivers who pursue a more aggressive driving style will be overrepresented in the CRE selection, since they will cross the envelope boundaries more often than drivers who prefer larger margins in general. Of course, if one assumes that aggressive driving is the main contributing factor to traffic accidents then this is OK. If one does not however, then there will be a selection bias in the CRE identification process that has to be dealt with somehow.

For FOT studies, this is actually quite feasible because CRE rates and dynamics with and without ADAS can be compared on an individual basis; thus taking out the effect of individual driving style. In a NDS-oriented study however, there will simply be more CREs for those driving with smaller margins, and some other means of dealing with that have to be identified.

2.5 Approach 4 - Driving history based identification

A fourth approach towards CRE detection is to look for unusual events in a driving history perspective, either on the individual or the group level. The idea is that unusual events in a person's or group's driving history are unusual precisely because the drivers try to avoid such events. Hence they can be said to represent situations which the driver is unwilling to enter, presumably because at least a certain portion of them would lead to potentially unsafe situations.

The advantage of this approach is that it will find the most unusual events that occurred during the study for each person or group (depending on the setup), and it is not unreasonable to assume that those rare events were ones which the drivers would prefer to avoid happening in the future. The corresponding disadvantage is of course that those events may be special for other reasons than being safety-critical, so even if drivers in a statistical sense try to avoid them, they may have little or no connection to traffic safety.

3. Coupling CREs and crash risk – the CRE causation question

As already discussed in the introduction, due to the very low risk of actual crashes occurring, NDS and FOT data sets typically contain only a few real crashes. Hence, the use of surrogate

events like non-crash CREs usually is motivated by referring to some version of Heinrich's classical triangle from his pioneering work on industrial safety in the 1930's (Heinrich 1931). Often, they cite Heinrich's proposed 600 to 1 ratio between incidents and accidents as a reason why incident studies may be a faster way to learn about problems that need solving.

However, what is often less clear in published studies is an exact description of how they connect the CREs they use to real crashes. Questions like whether the causation mechanisms are the same for the studied incidents as the real accidents, and if so, which data types must be collected to verify that the "right" CREs have been found is usually missing.

In all four approaches to CRE definition outlined above, there are certain underlying assumptions regarding this coupling between the CREs and real crashes. For example, to go with a driver response based CRE identification, it has to be assumed that kinematically drastic driver responses are predictive of crash involvement, e.g. that hard decelerations are the "bottom" of an iceberg where lead vehicle crashes are the top.

If instead a contextual approach is used, it has to be assumed that small margins are predictive of crash involvement, i.e. that small margin situations are the items at the bottom of the iceberg, where "no margin" situations constitute the top (i.e. actual crashes). Or if a driving history approach is choosen, it has to be assumed that rarely occurring combinations of kinematic (and other) values are predictive of crash involvement.

In other words, the approaches represent very different views of the coupling between CREs and crash risk. Which one (if any) is more appropriate remains to be determined on a project by project basis, as they all have different consequences for how the project findings can be extrapolated to the rest of the driving population.

4. What do the identified CREs represent?

The approaches described above also have different implications in terms of what the selected CREs represent. In the driver response based approach, CREs are identified based on how each driver evaluates the situation. For example, while one driver may brake hard at a certain time-to-collision threshold, another driver might not brake at all when at that time-to-collision value. Hence a selection of CREs based on this approach will include the first event but not the second (since the driver did not respond, it is by definition not an event).

Driver response based CRE selections will therefore reflect the normal variability in any driver population in terms of driving style, risk perception and capacity to respond. It follows that representative selection of drivers becomes a key issue when using a driver response centred CRE selection approach.

The function based CRE selection approach does present an interesting conundrum in terms of what the identified CREs really represent. Simply put, due to sensing and/or threat detection algorithm limitations, the function may not capture all the critical situations you think it should. Hence it would be preferable if you would could write and run your own threat detection algorithm on the collected data in order to avoid what a function limitation based CRE selection bias.

However, this is easier said than done. The developers of the function under FOT assessment are more often than not clever people who have worked on the function for several years. This means that the alternative threat detection algorithm you develop has to beat their algorithm, while still being based on the same sensor data. Unless you're very lucky or a genius, it might be unwise to rely on being able to do that.

Another option here is of course to add additional sensing capability before running the FOT, i.e. sensors which have much better performance than those used by the function being assessed, and then run a threat detection algorithm on that sensor data once collected. However, the additional equipment cost might be quite large, so you really have to believe there exists a class of important CREs outside the evaluated functions current detection capability. Moreover, it also must be necessary for the project to find them. For example, if you're tasked with defining the remaining safety potential in terms of avoiding rear end crashes given a particular FCW function, then this makes sense. Give this some thought before adding additional sensing to the project cost.

In a driving context approach, CREs are identified independently of how a particular driver responds to the situation. This means all drivers are equally covered by the conflict definition, independently of their capacity or willingness to respond. It also means that drivers with a more aggressive driving style will be overrepresented in terms of how many events they contribute to the list of CREs selected for analysis. They simply end up in small margin situations more often.

Now, if you assume that small margins in and of themselves are predictive of crash involvement, it does not matter that some drivers contribute with more CREs than others, since by this logic, these drivers do have a higher crash risk. On the other hand, if you're dubious about this assumption, then this approach might not be for you to begin with.

The advantage of a driving history based CRE selection approach is that one can tailor the CRE selection to each participating individual. Drivers all respond in some way when a potentially dangerous situation is sprung on them; but driver responses can be expected to vary even in complete surprise situations. For example, while 0.7G might be a normal deceleration level for one driver, another might never go above 0.65G regardless of how critical the situation is. Thus, if a 0.65G deceleration occurs only once in a person's full driving history, one might draw the conclusion that it was a special event for that driver and that it warrants further examination.

A prerequisite for the driving history based approach is that basic driver behaviour is fairly stable, i.e. that the distributions are not too "flat". For example, if you drive all over the lane rather than keeping fairly well to the lane centre, the tail of your lane position distribution will include both the situations when you ended up there unintentionally and would have preferred to be elsewhere, and the situations where you just ended up there and it didn't matter. If that is the case, the search algorithm will find most events simply due to a lot of inherent variability in the data, not because the driving situations as such have special, crash relevant, properties.

As an example, a technique called Gaussian Mixture Models (GMM - for an example from another domain, see Reynolds, 1995) was tested in the euroFOT project. Gaussian

distributions were fitted to a number of logged vehicle parameters such as brake pressure and pedal positions. Rare situations were then identified by looking for combinations of unusual values; i.e. where the values for several parameters were at the tail of their respective distributions simultaneously.

The GMM approach was found to work much better for truck drivers than for car drivers. For one, truck drivers drive more consistently due to spending so much more time behind the wheel. Furthermore, they also have a much smaller kinematic space to drift around in due to the size and mass of their vehicles. In other words, end points in their distributions therefore are more likely to represent truly unwanted values rather than random occurrences due to normal variability. Hence this approach might be more applicable for professional drivers who show less variability over time, rather than lay people who drive much less.

Obviously, any CREs identified will be more credible if they somehow link to underlying mechanisms related to the driver's control, e.g. over the vehicle, over positioning relative to other traffic participants, etc.

One recent approach that illustrates this is the work by Gordon et al (2011). They assumed that single vehicle crash risk depends on the driver's lateral control. The study assumed that the underlying mechanisms leading to such crashes are the same as those that create variations in normal driving—especially those involving "disturbed" lane-keeping control. The road-departure crash problem was formulated according to the following set of hypotheses:

- Single-vehicle road-departure crashes occur only under conditions of disturbed control
- Naturalistic Driving Data (NDD) contain measurable episodes of disturbed control
- Crash surrogates exist and are based on a combination of objective measures of disturbed control, highway geometric factors, and off-highway factors
- Crash surrogates can be related to actual crashes

The study compared lane deviation, LDW and instances of very short distance to the road edge as crash surrogates, and found the latter to be the best, with LDWs as the second best and variability within the lane to be the least successful predictor. In other words, in this study the empirical findings were predicted from theory and also followed the theory.

5. Removing false positives

Since there are no perfect CRE identifiers, any event search in a NDS/FOT database is likely to come up with a mix of true positives (i.e. actually relevant events) and false positives (i.e. events, while captured by the search algorithm, turn out not to be coupled to the proposed accident causation mechanism after all). An example would be a search algorithm that identifies hard braking events, but which cannot in and of itself distinguish if there is a lead vehicle in front or not.

Therefore, after a set of what could be called CRE candidates have been identified in any NDS/FOT data set, comes the very important step of weeding out the false positives from the

true positives. There are essentially three ways to do this; manual (visual) inspection, filtering based on some logic, or a combination.

A seemingly straightforward way to remove false positives from among the event candidates. process is manual visual inspection. Researchers review each event based on the video recorded and other available data, and then decide whether the event is truly crash relevant or not.

One advantage of this approach is that variables that might be impossible to capture numerically can be weighed in, such as how startled or scared the driver seems to be. On the other hand, assessing a level of criticality from comparatively low resolution videos is often difficult; video is not reality after all.

From this also follows that the assessment will vary between researchers. Inter-rater reliability thus becomes an important factor. Given that a reliability round 80% is considered a good number in inter-rater reliability studies, one realizes that if one of every five events is potentially misclassified, the statistics used in the final analysis have to compensate for that (if possible).

Another way to go about it is by applying one or more conditional filters. This means that the initial group of CREs is screened based on whether the CREs fulfil certain additional criteria. In the lane departure example, in addition to having inadvertently left the lane, one might remove all CRE candidates where the driver did not respond to the situation by steering or braking within three seconds, on the assumption that if the driver fails to respond the situation is not critical at all (unless there is a subsequent crash of course).

Another example is the filtering applied by Benmimoun et al in euroFOT (Benmimoun, Fahrenkrog et al. 2011), where various combinations of speed dependent yaw and deceleration filters are used to identify lateral conflicts.

The search for a good CRE definition is often an iterative process, where the selection criteria are successively refined to suit the study purposes. As an example of such an iterative process, Fitch et al (Fitch, Rakha et al. 2008) used an iterative process where the initial pool of CRE candidates was screened using one new filter at a time. After each filtering, visual inspection of a randomly sampled portion of the remaining CRE candidates was carried out, and this process continued until the desired ratio between true and false positives was achieved.

5.1 Which way to go?

It is not obvious which of these approaches to filtering is the best way to go. Please also note that no formal restrictions apply, in the sense that if you select approach you must stick to that approach only. Some projects have used combinations of the above approaches to find events, and this is perfectly fine.

What is clear however is that different CRE definitions may lead to very different results, even when used on the same data. In the above study (Fitch, Rakha et al. 2008), it is described that another project (Hanowski, Blanco et al. 2008) used a kinematic threshold +

visual manual CRE candidate selection on the same data set as Fitch et al were using. The opportunity therefore arose to compare CRE selections between the two projects. Interestingly, while both projects found hundreds of what they judged to be relevant CREs, only 7 of the 596 conflicts found by Hanowski et al study were identified in the Fitch et al study.

This shows that if anything, the acronym WYLFIWYF (What-You-Look-For-Is-What-You-Find) coined by Erik Hollnagel (Hollnagel 2004)) holds for NDS/FOT analysis and CRE identification as well. It also shows that the most important requirement here is that you explicitly describe why you did choose the CRE definition you end up with in the end, i.e. how that particular CRE type is linked to the crash type(s) you are trying to understand and prevent.

6. The problem of Baseline selection

When analysing CREs, whether it is their relative frequency or some particular aspect of driver behaviour during the CREs, one always has to define a baseline, i.e. a comparison situation. This is one of the philosophically more challenging aspects of both NDS and FOT studies.

6.1 Naturalistic Driving Study baselines

In NDS studies, a key topic of interest is whether any particular driver behaviours are overly represented in CREs, i.e. whether they can be viewed as crash contributing factors. To find this out, one has to have a set of relevant comparison situations, i.e. a baseline, to compare the CREs to.

There are two principal ways to select baseline events (non-CRE events). One is to randomly sample a number of baseline events from all data that is not part of the CREs. The other is to tailor the baseline selection to the participants. For example, if a CRE occurs on a highway in the afternoon when it rains, then one tries to find a baseline event that also occurs on a highway in the afternoon when it rains. Naturally, this is easier for conditions that occur more often. Finding one or more matching baseline events for a CRE that occurred while commuting to work is easier than finding a match for a CRE that occurred on a vacation road trip.

More formally, these two approaches represent two types of experimental design (case-cohort vs. case-crossover designs) Rather than going into what can be expected given that either approach is used, the reader is referred to the in-depth comparison of the strengths and weaknesses of each approach when used on the 100 car dataset (Guo and Hankey 2009).

6.2 FOT baselines

In a Field Operational Test, the baseline selection problem is slightly different compared to an NDS. The example of Adaptive Cruise Control (ACC) illustrates this point very well. It is relatively straightforward to say that the relative frequency of CREs that occur while ACC is in use constitutes the treatment data. However, what should that be compared to?

One way to resolve this is to focus on the adaptive part of the cruise control functionality. This means one would compare the relative CRE frequency when ACC is being used to the CRE frequency when regular cruise control is being used. Another way is to focus only on CREs that occur in car following situations. Here, the baseline would be all CREs that occur when there is a lead vehicle present and ACC is not engaged.

More examples can easily be constructed, but these two suffice to illustrate that baseline selection inherently depends on the researchers perception of which the relevant accident causation mechanisms are.

As an example of how complex this may get, consider the following approach that finally was settled on and used in euroFOT:

The approach adopted in euroFOT was to exclude all treatment data in which ACC was OFF. This had the advantage that it ensures 100% usage in the treatment portion of the analysed data. However, it caused another problem: the driver selection bias. Since ACC usage is self-paced by drivers (i.e. the drivers themselves decide where and when to use it), the baseline data should principally be selected to include only driving where the drivers would have opted to use ACC, had it been available. Defining filters for selecting data according to this principle is very complicated and might be impossible, as it involves second guessing driver behaviour. Instead, a set of filters thought to approximate the perfect baseline data selection was used. These filters, which were applied to both baseline and treatment data, and their effect in equalizing the driving conditions, are described in the following:

Car following filter: In terms of safety, ACC is mainly targeted towards reducing the number and severity of rear-end crashes. Hence, only data from driving when a lead vehicle was present in front of the equipped vehicle was included.

Posted speed filter: data in speed limits in which ACC was not used very often (usage below 25%) were discarded.

Vehicle speed filter: When approaching roundabouts and larger intersections, drivers normally brake. As this automatically disengages ACC and hence inevitably excludes this data from the treatment set (ACC on), it had to be removed from baseline as well. A simple way of tackling this issue is to set a limit on minimum vehicle speed. Drivers typically enter roundabouts and larger intersection with speed equal to or below 50km/h, so by setting a minimum vehicle speed of 50km/h, most junction and roundabout driving was removed from both baseline and treatment periods.

5 seconds wait-and-see filter: as mentioned above, ACC disengages at braking. This means that harsh braking and critical time-gap events that happen right after ACC disengages would not be included in treatment if the ACC ON filter was strictly applied in treatment. To compensate for this, treatment data was selected to include five additional seconds each time ACC disengaged, to make sure these this type of events were coupled to ACC usage and not excluded from the treatment data.

7. Conclusions

Unfortunately, there is yet no formal CRE recipe available that allows just for doing and not thinking. Thus, as hopefully has become clear in the above discussion, setting up CREs and the corresponding baseline events requires thinking about crash causation mechanisms, driver selection principles, what actually constitutes risk-free driving, and many other things.

The aim of this chapter is to describe common approaches to CRE selection on a higher, grouped level, and list some of the pros and cons for each CRE group. **The whole point of this exercise is to support and motivate you and your project to make an informed and conscious decision as to which approach to CRE selection best will fulfil your project goals, and then set up your CREs correspondingly.**

However, lets not forget that progress is being made in many concurrent projects. Hopefully, we will see convergence toward common CRE definitions in the next few years. Also, while most work until now has focused on vehicle-to-vehicle conflicts, there are now also projects underway that will broaden the scope by looking at conflicts with vulnerable road users (pedestrians, bicyclists, etc.) as well as Powered Two-Wheelers. While each of these conflict objects represent new challenges when it comes to defining the actual crash causation mechanisms involved, as well as which surrogate events could be used as indicators of accident occurrence likelihood, they also drive a good portion of creative thinking on the subject, which is a good thing.

8. References

Benmimoun, M., F. Fahrenkrog, et al. (2011). Incident Detection Based on Vehicle Can-Data within the Large Scale Field Operational Test “euroFOT”. 22nd Enhanced Safety of Vehicles Conference. Washington, D.C.

Dingus, T. A., S. G. Klauer, et al. (2006). The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment. Cambridge, U.S. Department of Transportation, John A. Volpe National Transportation Systems Center.

Fitch, G. M., H. A. Rakha, et al. (2008). Safety Benefit Evaluation of a Forward Collision Warning System, US DOT, NHTSA, Report No DOT HS 810 910, Washington, D.C..

Tim J. Gordon; Lidia P. Kostyniuk; Paul E. Green; Michelle A. Barnes; Daniel Blower; Adam D. Blankespoor; Scott E. Bogard, Analysis of crash rates and surrogate events: Unified approach (2011) Transportation Research Record. 2011;(2237):1-9.

Guo, F. and J. Hankey (2009). Modeling 100-Car Safety Events: A Case-Based Approach for Analyzing Naturalistic Driving Data, The National Surface Transportation Safety Center for Excellence.

Hanowski, R. J., M. Blanco, et al. (2008). The Drowsy Driver Warning System Field Operational Test: Data Collection Methods, Virginia Tech Transportation Institute.

Heinrich, H. W. (1931). Industrial accident prevention: a scientific approach. New York, McGraw-Hill.

Hollnagel, E. (2004). Barriers and accident prevention. Burlington, VT, Ashgate.

Kiefer, R., M. A. Shulman, et al. (2003). Forward collision warning requirements project: refining the CAMP crash alert timing approach by examining "last-second" braking and lane change maneuvers under various kinematic conditions.

LeBlanc, D., T. J. Gordon, et al. (2006). "Road departure crash warning system field operational test: methodology and results. Volume 1: technical report." 307 p.

Najm, W. G., M. D. Stearns, et al. (2006). Evaluation of an Automotive Rear-End Collision Avoidance System, Research and Innovative Technology Administration, Volpe National Transportation Systems Center , National Highway Traffic Safety Administration: 425p.

Reynolds, D., Rose, R.C.,(1995) Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. IEEE Transactions on Speech and Audio Processing, Vol 3., No 1.